UNIVERSITY OF CALIFORNIA,
IRVINE

Ecological and evolutionary consequences of microbial community responses to environmental change

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Sarai S. Finks

Dissertation Committee
Professor Jennifer B.H. Martiny, Chair
Associate Professor Katrine L. Whiteson
Professor Adam C. Martiny

2022

DEDICATION

Para

mi madre y hermana,

Nancy Carol Arreola Finks y Carol Lyn Finks

a quienes yo amo mucho por toda la ayuda que me han dado

Y para

mis abuelos,

Efraine Cardona Arreola y Carolina Garza Arreola

muchas gracias por el legado que han dejado en los Estados Unidos

Y, finalmente

Para

mi persona favorita,

Jack Liam Finks

gracias por todas las risas que me has dado mi niño

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to convey my deepest appreciation and gratitude for my committee chair, mentor, and friend, Professor Jennifer B.H. Martiny, who has shown me what it is to be a great scientist and human being. Through her continuing encouragement and belief in me to accomplish what I set out to do during my PhD, I have learned to trust myself. Jen nurtured a safe environment that also challenged me to think about my science in different ways. Without her creative genius and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Associate Professor Katrine L. Whiteson and Professor Adam C. Martiny, whose thoughtful questions and perspectives challenged and encouraged me to find meaning in my science. I thank them for their continued support and sending relevant articles my way that helped me in answering the questions I set out to address. I thank all my committee for their patience, especially, during the times where articulating my science was convoluted and perhaps myopic. Truly, I have learned so much from you and I am better prepared for the next steps in my career.

In addition, many thanks to Professor and Associate Dean for Graduate Studies, R. Michael Mulligan and Professor Kathleen K. Treseder for their support and intervention during a very transitional time of my dissertation, that without their help would not have been possible.

I also acknowledge and thank the funding sources who made my dissertation research possible. Fellowship support was provided by the National Science Foundation Graduate Student Fellowship Program, the UCI Chancellor's Club Fund for Excellence Fellowship, National Institutes of Health Grant NIH 4R25GM055246-20, National Science Foundation Grant DEB-1457160, and the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research grants, DE-SC0016410 and DE-SC0020382.

Chapter 1 was published with the permission of the University California Press. The text of this chapter is a reprint of the material as it appears in the journal *Elementa: Science of the Anthropocene*.

VITA

Sarai S. Finks

Research Interests

Microbial Ecology and Evolution; Global Change Ecology; Plant(host-)Microbe interactions; Horizontal Gene Transfer; Mobile Genetic Elements; Evolution of Antibiotic Resistance.

Education

2015 – 2022	Doctor of Philosophy (Ph.D.) Biological Sciences	*University of California – Irvine (UCI)*
2015 – 2019	Master of Science (M.S.) Biological Sciences	*University of California – Irvine*
2010 – 2012	Bachelor of Science (B.S.) Biological Sciences	*University of California – Irvine*
2007 – 2010	Saddleback College	

Professional Appointments

2022 – *Postdoctoral Researcher*
Department of Entomology and Biology at Pennsylvania State University
Principle Investigator: Seth Bordenstein

2018 – 2022 *Research Assistant*
Department of Ecology and Evolutionary Biology, UCI
Principle Investigator: Jennifer B.H. Martiny

NSF DEB-1457160, DOE DE-SC0016410 and DE-SC0020382.
Collaborative Research: Investigating Microbial community responses and adaptation to environmental change and simulated global changes.

2015 – 2017 *Research Assistant*
Department of Ecology and Evolutionary Biology, UCI
Principle Investigators: Francisco J. Ayala and Luis Mota-Bravo

NIH 4R25GM055246-20. Collaborative Research: Investigating the evolution of antibiotic resistant bacteria from aquatic environments.

2012 – 2015 *Laboratory Assistant II*
Minority Sciences Program, UCI

Publications

3 **Finks, S.S.,** and J.B.H. Martiny. 2022. Plasmid-encoded traits vary across environments. *mBio* (In Revision).

2 Scales, N., A.B. Chase, **S.S. Finks**, A. Malik, C. Weihe, S.D. Allison, A.C. Martiny, J.B.H. Martiny. 2022. Differential Response of Bacterial Microdiversity to Simulated Global Change. *Applied and Environmental Microbiology*.

1 **Finks, S.S.,** S. Kimball, C. Weihe, A.C. Martiny, K. Treseder, S.D. Allison, and J.B.H. Martiny. 2021. The responses of microbial decomposers to a decade of simulated global changes depend on the plant community. *Elementa: Science of the Anthropocene*.

Publications in Preparation

Finks, S.S., P. Moudgalya^**,** C. Weihe, and J.B.H. Martiny. Plasmids drive trait variation within the soil taxon *Curtobacterium.*

^ High School Mentee

Awards and Honors

2021	UCI Chancellor's Club Fund for Excellence Fellowship ($12,000)
2020	UCI EEB Antiracism, Diversity, Equity, and Inclusion Fellowship ($1,500)
2017	National Science Foundation Graduate Research Fellowship ($102,000)
2016	Honorable Mention National Science Foundation Graduate Research Fellowship
2010 &2012	Edison STEM Transfer Scholarship ($15,000)

Presentations

Finks, S.S.*, P. Moudgalya^, C. Weihe, and J.B.H. Martiny. 2022. Mobile genetic elements drive trait variation within the soil taxon *Curtobacterium.* ISME18: 18th International Symposium on Microbial Ecology. Lausanne, Switzerland. Poster Presentation.

Finks, S.S.*and J.B.H. Martiny. 2021. The Plasmidome across Environments: a Nexus of Trait Diversity in Bacterial Communities. ASM & FEMS World Microbe Forum. Virtual. Invited Talk.

Finks, S.S.*, S. Kimball, C. Weihe*, A.C. Martiny, K. Treseder, S.D. Allison, and J.B.H. Martiny. 2020. The effect of drought on bacterial community composition and plant litter decomposition. DOE Genomic Sciences Program Annual PI Meeting. Washington, DC. Poster Presentation.

Finks, S.S.*, M. De la Cruz, L. Mota-Bravo, F.J. Ayala. 2017. Extensively antibiotic-resistant Escherichia coli harboring multidrug-resistant conferring conjugational plasmid found in lake water. Keystone Symposia on Molecular and Cellular Biology (Antimicrobials and Resistance: Opportunities and Challenges T4). Santa Fe, NM. Poster Presentation.

* Presenting author ^ High School Mentee

Advanced Coursework and Workshops

UCI -Coursework

2017 Ethical Conduct of Research
2017 Proposals and Grant Writing
2016 Comparative Physiology
2016 Special Topics in Ecology
2016 Special Topics in Evolution
2016 Special Epidemiological Topics: Bioinformatics
2015 Quantitative Statistical Methods in Ecology and Evolution

UCI – Workshops

2019 Reclaiming STEM west coast
2019 The Center for Microbiome Science & Ridge2Reef: Microbiomes and Global Change
2019 Programming with Python
2019 Introduction to Machine Learning in Python
2018 Microbiome data analysis
2018 Becoming an Effective
2016 Phylogenetic analysis using RevBayes
2016 Introduction to Unix/Linux
2015 IANAS-UNESCO-IHP: Water Qualities in the Americas

Teaching

2022 An Introduction to Metagenomics (UCI – The Center for Microbiome Science). Assisted in curriculum development and teaching analysis of metagenomic sequencing data of microbial communities using MetaPhlAn, HUMAnN, and R.

2019 Introduction to R-Based Analysis of Microbiome Data (UCI – The Center for Microbiome Science). Assisted in teaching analysis of microbial community data using R and QIIME2.

2018 Human Biology (UCI – Department of Neurobiology, N120C). Reviewed students weekly writing assignments and provided constructive feedback on assignment, and graded quizzes and tests.

Mentoring

2019 – 2021 Pranav Moudgalya (High School Mentee)
Advised through the Southern California Junior Academy of Sciences Program,
Project: Plasmids of *Curtobacterium*: a nexus of CAZyme diversity,
Currently an undergraduate student at Harvard University.

2020 -2019 Harsha Rao (UCI Masters of Business Technology Mentee)
Project: Biogeography of plasmids in the soil taxon *Curtobacterium,*
Currently a Product Manager at Bio-Rad Laboratories.

2017 -2017 Freddie Adame (UCI Biological Sciences Undergraduate Mentee)

Advised through UCI Minority Science Program
Project: Multidrug-Resistant Plasmid Found in Environmental *Escherichia coli* Harbors bla_{TEM-1b} and $bla_{CTX-M-15}$ Extended-Spectrum β-lactamase,
Currently a Medical Assistant at Total Family Care and Walk-in Clinic, LLC

2017 - 2017 Mira China-Ly Rice (UCI Biological Sciences Undergraduate Mentee)
Project: Multidrug-Resistant Plasmids of *Enterobacteriaceae* from Aquatic Environments

Outreach

2021 Member of the Antiracism, Diversity, Equity, and Inclusion Council for the EEB Department at UCI. The council is tasked with promoting ARDEI principles in departmental policies and practices.

2021 – 2018 Member of the UCI chapter of SACNAS, an inclusive organization dedicated to fostering the success of Chicanos/Hispanics and Native Americans, from college students to professionals.

2020 Panelist for What's next: Is grad school for me? Answered undergraduate URM students questions to help demystify the graduate application process for first generation college considering careers in STEM.

2020 Mentor-Judge for SACNAS National Diversity in STEM Virtual Conference. The conference serves to equip, empower, and energize underrepresented minority participants for their academic and professional paths in STEM.

2020 Organizer of the Research Opportunities for Transfer Students Panel. Where Graduate students shared research opportunities with undergraduate students participating in the Saddleback College's Chicano, Latino, Access, Success & Empowerment transfer mentor program.

2019 – 2017 NSF GRFP writing workshop (UCI – School of Biological Sciences). Reviewed graduate student personal and research statements
and provided constructive feedback.

2019 -2018 Board member of the Orange County Society for Conservation Biology
A local chapter of the Society for Conservation Biology aimed at sharing conservations issues relevant to the Orange County with the public.

2017 Initiative for Maximizing Student Development Research Program (UCI – Minority Science Program). Taught basic laboratory skills to underrepresented undergraduate minority students, CLSI antimicrobial testing and basic principles in antimicrobial resistance.

2017 & 2015 Poster Judge for the MSP Summer Research Symposium

2016 Tech-Trek Summer Camp Teacher (American Association of University Women). Taught middle-school aged womxn interested in STEM careers how to identify and characterize antibiotic resistant bacteria.

2015 BRIDGES to Baccalaureates Research Program (UCI – Minority Science Program) Taught basic laboratory skills to underrepresented minority students from community colleges in preparation for summer research in labs across UCI.

ABSTRACT OF THE DISSERTATION

Ecological and evolutionary consequences of microbial community responses to environmental change

by

Sarai S. Finks

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2022

Professor Jennifer B.H. Martiny, Chair

Global changes such as increased frequency of fire, drought, and nitrogen deposition, perturb microorganisms and the higher trophic life forms they support. Microorganisms play key roles in carbon and nutrient cycling, which are important to agriculture and ecosystem health. Although microorganisms are pivotal in an ecosystem's response to environmental changes, little is known about how abundant and diverse microbial communities adapt to such changes. The overarching aim of my thesis is to investigate how bacterial communities respond to global change and in particular, their ability to quickly adapt to environmental perturbations.

I first investigated how microbial responses to global changes are influenced by interactions with plant communities using the Loma Ridge Global Change Experiment, a decade-long experiment that manipulates rainfall and nitrogen levels across two adjacent ecosystems (Chapter 1). My findings underscore the importance of plant–microbe interactions when considering the transferability of the results of global change experiments across ecosystems. Next, I investigated traits found on plasmids, a type of mobile genetic element (MGE) that can facilitate rapid evolution in bacteria. I asked what are the ecologically-relevant plasmid genes that may serve as reservoirs of environmental-adaptive traits in bacteria (Chapter 2). The

findings of this chapter suggest that plasmid traits may contribute to host adaptation in environmental microbiomes. Lastly, I extended this work to a cosmopolitan soil taxon, *Curtobacterium*, an abundant genus of bacteria in southern California ecosystems. This taxon shows marked shifts in relative abundance in response to simulated drought and is amenable to culturing, providing a tractable system for investigating both genotypic and phenotypic characteristics of this organism. Previous experiments have shown *Curtobacterium* rapidly evolve via *de novo* mutations in response to environmental changes. I asked what MGE and associated traits are found in *Curtobacterium*, and determined whether MGE and traits showed any environment- versus clade- specific genomic signatures (Chapter 3). The findings of this chapter highlight the potential of traits found on plasmids to be mobilized within the bacterial communities where these *Curtobacterium* were isolated. Overall, my thesis work highlights the importance for considering the intersection of evolution and ecology in understanding how microbial communities adapt to environmental changes.

INTRODUCTION

"The first step toward genetic novelty is the origination or acquisition of genomic variation"
- Kirchberger, Schmidt, and Ochman[1]

The total number of microorganisms in terrestrial environments is estimated to be approximately 10^{29} (Flemming and Wuertz, 2019), and soil microbial communities are among the most diverse. These communities influence ecosystem processes through resource assimilation into biomass and regulating carbon inputs via decomposition of senescent plant matter (Schimel and Schaeffer, 2012; Gleixner, 2013; Kuypers et al., 2018). Global changes such as drought and increased atmospheric nitrogen impact soil microbiome diversity, abundance, and ecosystem functioning like decomposition of plant material (Alster et al., 2013; Matulich and Martiny, 2015; Martiny et al., 2017; Glassman et al., 2018). Since soil carbon and nitrogen stocks reflect a dynamic balance between microbial decomposition of organic carbon and assimilated nitrogen (Mooshammer et al., 2014), understanding the mechanisms underlying how microorganisms respond and adapt to environmental disturbances is key in determining the long-term ecological and biogeochemical consequences of anthropogenic activities and climate change (Cavicchioli et al., 2019).

Bacteria can adapt quickly to perturbations due to their relatively fast growth rates, high mutation rates, and horizontal gene transfer (HGT) (Thomas and Nielsen, 2005). Thus, it is often assumed that microbial functioning in natural ecosystems will be highly resistant to changes. Indeed, rapid adaptation in bacteria is showcased by the global spread of antibiotic resistance in healthcare-associated communities where HGT of resistance genes is facilitated by MGE such as

[1] Kirchberger PC, Schmidt ML, Ochman H. 2020. The Ingenuity of Bacterial Genomes. *Annu Rev Microbiol* **74**(1): 815–834. Annual Reviews. doi: 10.1146/annurev-micro-020518-115822

plasmids (Johnson et al., 2007; Carattoli, 2009; Nordmann et al., 2011; Carattoli, 2013; Hu et al., 2014; Li et al., 2015). Mechanisms of HGT (e.g., conjugation, transduction, and transformation) have been well studied in laboratory settings using single bacterial strains, and much of the trait variation in bacteria is thought to be attributed to HGT (Wollman and Jacob, 1956; Curtiss III, 1969; Johnsborg et al., 2007; Johnston et al., 2014; Kirchberger et al., 2020; Schneider, 2021). However, there is a surprising disconnect between what is known about HGT in the lab and our understanding of these processes in natural environments (Brito, 2021). This contributes significant uncertainty in our ability to predict how bacteria living in complex natural communities will respond to environmental change.

Since microbes often live amongst other microbes, and macroorganisms such as plants, species niches in these complex communities are constantly adjusting in response to changing environmental conditions (Dolph et al., 1985; Rainey and Travisano, 1998; Hairston Jr et al., 2005; Lawrence et al., 2012; WM J. et al., 2013; Stuart et al., 2014). For example, in southern California, wildfires and severe drought not only impact soil microorganisms but plant communities as well, which in turn can indirectly impact microbes. At the microbial level, physical stressors like osmotic and heat-shock may create the ideal conditions for HGT between interacting microbes, which may result in fitness advantages on members of the soil microbiome. Since climate change responses will likely depend on the traits involved (Allison, 2012; Krause et al., 2014; H. et al., 2015; Ashish A Malik et al., 2020), the exchange of traits in microorganisms under environmental pressure may contribute to the resilience of ecosystem functioning and soil biogeochemical processes as these ecosystems are increasingly perturbed by climate change (Dai, 2011; Sheik et al., 2011; Gao et al., 2011; Kinugasa et al., 2012; Schmidt et al., 2018; Griffin-Nolan et al., 2019; I. et al., 2022). Understanding which biotic interactions

impact microbial responses whether through macroscale interactions or microbial interactions via the horizontal exchange of traits will help inform predictions about how microbial communities change and ecosystems function with projected hotter, drier climates.

CHAPTER 1

Microbial community response to a decade of simulated global changes depends on the plant community

ABSTRACT

Global changes such as increased drought and atmospheric nitrogen deposition perturb both the microbial and plant communities that mediate terrestrial ecosystem functioning. However, few studies consider how microbial responses to global changes may be influenced by interactions with plant communities. To begin to address the role of microbial-plant interactions, we tested the hypothesis that the response of microbial communities to global change depends on the plant community. We characterized bacterial and fungal communities from 395 plant litter samples taken from the Loma Ridge Global Change Experiment, a decade-long global change experiment in southern California that manipulates rainfall and nitrogen levels across two adjacent ecosystems, a grassland and a coastal sage scrubland. The differences in bacterial and fungal composition between ecosystems paralleled distinctions in plant community composition. In addition to the direct main effects, the global change treatments altered microbial composition in an ecosystem-dependent manner, in support of our hypothesis. The interaction between the drought treatment and ecosystem explained nearly 5% of the variation in bacterial community composition, similar to the variation explained by the ecosystem-independent effects of drought. Unexpectedly, we found that the main effect of drought was approximately four times as strong on bacterial composition as that of nitrogen addition, which did not alter fungal or plant composition. Overall, the findings underscore the importance of considering plant-microbe interactions when considering the transferability of the results of global change experiments across ecosystems.

INTRODUCTION

Ongoing human-driven global changes are altering species distributions and interspecific interactions (Parmesan, 2006; Rosenzweig et al., 2008). Many field experiments have considered the responses of plant and/or microbial communities to simulated global changes (Stylinski and Allen, 1999; Vilà et al., 2003; Allison et al., 2013; Martiny et al., 2017). These studies aim to predict how global changes such as altered precipitation, nitrogen availability, CO_2 concentration, and temperature affect community composition and ecosystem processes (Stylinski and Allen, 1999; Cione et al., 2002; Vilà et al., 2003; Allison and Martiny, 2008; Cruz-Martinez et al., 2009; Gaertner et al., 2009; Castro et al., 2010; Gutknecht et al., 2012). However, fewer studies consider how the impact of a global change treatment on microbial composition is influenced by interactions with the plant community (Classen et al., 2015; Sayer et al., 2017). Such interactions will influence the transferability of the results of global change experiments as plant community composition, among other factors, vary across ecosystems.

Here, we focus on one half of plant-microbe interactions – specifically, the ways in which plants may influence microorganisms. One way in which plants influence microorganisms is through decomposition. Bacteria and fungi are the primary decomposers of dead plant biomass, and this process regulates the amount of soil carbon exchanged with the atmosphere (Swift et al., 1979; Adair et al., 2008; Schimel and Schaeffer, 2012). More broadly, plant communities can influence microorganisms through plant species and tissue composition (influencing nutrients and secondary compounds), changes in the abiotic environment (plant architecture influencing canopy and moisture of the soil), and relationships with other organisms such as endophytes and herbivores (Wardle et al., 2006; Tintjer and Rudgers, 2006; Kara et al., 2008; Kominoski et al., 2009; Rodriguez et al., 2009; Chapman and Newman, 2010; Santonja et al., 2017; Graff et al., 2020). Through such associations, the impact of a global change treatment on plant-litter

microbial communities may depend on the plant community in at least three ways. First, initial differences in plant communities across ecosystems will select for different microbial communities. Thus, the microbial response to global change may be driven by taxa that are uniquely prominent in an ecosystem (Ashish A. Malik et al., 2020). Second, some microbial taxa may be more sensitive to global change depending on the plant resources available to them in an ecosystem (Wood et al., 2018; Ashish A Malik et al., 2020). Specifically, litter substrate quality may impact the ability of litter microbes to respond to stressful conditions (Ashish A. Malik et al., 2020). Finally, the extent to which the plant litter substrates and resources are altered by global change, and thereby indirectly influence litter microbial communities, will depend on the plant community (Aerts, 1997; Rouifed et al., 2010; Fernández-Alonso et al., 2018). For instance, drought typically reduces the germination of annual plants, ground cover, and primary productivity of arid grasslands, whereas plants with deeper root systems are less impacted (Le Houérou, 1996; Shinoda et al., 2010; Kinugasa et al., 2012).

The Loma Ridge Global Change Experiment (LRGCE) simulates the increased frequency of drought and the increased availability of nitrogen. Drought is an extreme climatic event that occurs in most climatic zones, and its frequency and severity are projected to increase, along with atmospheric nitrogen deposition (Mishra and Singh, 2010; Dai, 2011; IPCC, 2014). A unique feature of the LRGCE is that treatments are applied to two adjacent ecosystems (Figure S1A), a grassland and a coastal sage scrubland (CSS). Much is known about the plant and litter microbial communities at the LRGCE. Within the first five years of the experiment, both the grassland and CSS plant communities responded to drought and nitrogen addition (Potts et al., 2012; Kimball et al., 2014; Kimball et al., 2016). In the grassland, drought reduced non-native annual grass cover, while nitrogen addition reduced native grasses and increased non-native

annual grasses (Kimball et al., 2014; Kimball et al., 2016). In the CSS, drought reduced shrub cover and increased grass cover, and added nitrogen further reduced shrub cover and native grasses (Kimball et al., 2014). Bacterial and fungal community composition on surface plant litter also responded to these treatments in the grassland (Allison et al., 2013; Berlemont et al., 2014). Moreover, reciprocal transplant experiments within LRGCE revealed shifts in microbial community composition due to direct, abiotic effects of the global change treatments, and as indirect effects of drought on the grassland plant litter (Martiny et al., 2017). However, the response of the microbial communities within the CSS treatment plots at the Loma Ridge research site have not yet been investigated. Additionally, drought and added nitrogen treatments have been ongoing for more than a decade, thus it is important to assess the long-term effects of simulated drought and added nitrogen.

Here, we tested the hypothesis that the response of microbial communities to global change depends on the plant community using a decade-long global change experiment in southern California. To address our hypothesis, we ask: does the response of microbial communities to global change depend on the ecosystem (grassland versus CSS)? The two ecosystems are immediately adjacent to each other at our study site, without major differences in slope, aspect, soil type, or climate, and subtle bulk soil differences seem unlikely to affect the plant litter microorganisms on the soil surface. Thus, we presume that any differences in the treatment responses across ecosystems are likely due to differences in the plant communities, rather than other abiotic factors. As observed in the grassland within the first five years of treatments, we expect a treatment response would occur in both ecosystems, resulting in a significant main effect of the treatments. However, we also predict that the microbial community response will result in ecosystem-specific compositional shifts, as reflected in significant

treatment-by-ecosystem effects. Given that the surrounding abiotic conditions are similar in these adjacent ecosystems, such interactive effects provide evidence that microbial responses are dependent on the plant community.

MATERIALS AND METHODS

Field site, sample collection, and experimental design. The LRGCE was established in February 2007 and is located 5 km north of Irvine, California, USA (117.704°W, 33.742°N; 365 m elevation), on a sloping (< 10%) deep colluvial deposit from layers of sedimentary rock and soil mapped as Myford Sandy Loam (Potts et al., 2012; Kimball et al., 2014). The grassland plots are dominated by the native perennial grass *Stipa pulchra,* the annual grass genera *Avena*, *Bromus* and *Festuca*, and the annual forb genera *Erodium* and *Lupinus*. The CSS plots are dominated by *Artemisia californica, Salvia mellifera, Eriogonum fasciculatum,* and *Acmispon glaber* (Kimball et al., 2014). The climate is Mediterranean with an annual precipitation of 30 cm. The 'wet' season is typically from November to April while the 'dry' season is from May through October (Figure 1.1). Air temperatures are moderate in the wet season with an average high and low of 21.1°C and 7.1°C, respectively, and increase in the dry season with an average high and low of 27.6°C and 14.4°C, respectively (Tustin Irvine Ranch weather 1981 to 2010; Western Regional Climate Center[2].

Surface litter samples were collected at seven time points (approximately every three months) from August 2016 through March 2018, and from four replicate plots receiving four different treatments: control, drought, added nitrogen, and drought plus added nitrogen. Thus, up to 16 samples in each ecosystem were collected at each time point; however, after quality checks some samples were excluded for a total of 108 grassland and 111 CSS samples. The LRGCE

[2] https://wrcc.dri.edu/

implements a randomized split-plot design in both grassland (6.1 x 12.2 m) and CSS (18.29 x 12.19 m) sites, where the nitrogen treatment is nested within drought treatment plots. Drought control plots received ambient rainfall while drought plots were exposed to approximately a 50% reduction in rainfall, and either ambient or added nitrogen. Drought was simulated by covering rain shelters with manually retractable, clear, 6-mil polyethylene roofs before predicted large rain events and removed promptly afterwards (Figure 1.2B and C). Added nitrogen was applied at 60 kg N ha^{-1} $year^{-1}$ as fast release calcium nitrate, $CaNO_3$, and in two amounts per year, 20 kg prior to the first rains of the wet season and 40 kg in December coinciding with the start of the plant growing season (Potts et al., 2012; Kimball et al., 2014).

DNA isolation, PCR, and microbial community sequencing. DNA was extracted from ~ 0.05 g of senescent leaf material from treatment plots in grassland and CSS sites using ZymoBiomics DNA isolations kits (Zymo Research, Irvine, California, USA) and processed for 5 min of bead beating at maximum speed (6.0 m/s, FastPrep-24 High Speed Homogenizer, MP Biomedicals, Irvine, California, USA). To avoid batch effects, the plant litter samples were randomized prior to DNA extraction.

To characterize bacterial composition of the leaf litter communities, PCR amplification of the V4 and partial V5 region of the bacterial 16S rDNA (~ 411 bp) was carried out following the Earth Microbiome protocol (Lane et al., 1985; Caporaso et al., 2012; Parada et al., 2016). The barcoded forward primers contain the 5' Illumina adapter (AATGATACGGCGACCACCGAGATCTACACGCT), a unique 12-base error correcting Golay barcode, a pad (TATGGTAATT), a linker sequence (GT), and the 515fb primer (GTGYCAGCMGCCGCGGTAA); and reverse primers that contain the reverse complement of

the 3' Illumina adapter (CAAGCAGAAGACGGCATACGAGAT), a pad (AGTCAGCCAG), a linker sequence (GG), and the 926r primer (CCGTCAATTCCTTTRAGTTT).

Each 16S PCR reaction contained: 9.5 µL PCR grade water (Fisher Scientific, Hampton, New Hampshire, USA), 12.5 µL of 2x concentrated AccuStart II PCR ToughMix (Quanta bio, Beverly, Massachusetts, USA) for a final 1x concentration, 0.5 µL of 10 µM 926r primer (final concentration of 0.2 µM), 1 µL of 10 mg/mL bovine serum albumin (final concentration of 1 µg/mL; New England BioLabs, Ipswich, Massachusetts, USA), 0.5 µL of 10 µM barcoded 515f primers (final concentration of 0.2 µM), and 1 µL of genomic DNA. Reactions were held at 94 ℃ for 3 minutes to denature DNA, with amplification proceeding for 30 cycles at 94 ℃ for 45 seconds, 55 ℃ for 30 seconds, and 72 ℃ for 1 minute; followed by a final extension for 10 minutes at 72 ℃ to ensure complete amplification.

To determine fungal community composition, and improved accuracy of amplicon cluster detection and resolution during sequencing, a staggered primer design was used to amplify the internal transcribed spacer (ITS) region (~ 340 bp) of the 5.8S rRNA gene. The ITS primers used are as follows: ITS9f primer (AATGATACGGCGA CCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGAACGCA GCRAAIIGYG), along with a barcoded reverse ITS4 primer (CAAGCAGAAGACGGCATACGAGAT **********AGTCAGTCAGCCTCCTCCGCTTATTGATATGC), which contained the reverse complement of the 3' Illumina adapter, a unique 12-base barcode, and a pad with a linker sequence (Tremblay et al., 2015; Looby et al., 2016).

Each ITS PCR reaction contained: 9 µL PCR grade water (Fisher Scientific, Hampton, New Hampshire, USA), 12.5 µL of 2x concentrated AccuStart II PCR ToughMix (Quanta bio,

Beverly, Massachusetts, USA) for a final 1x concentration, 0.75 µL of 10 µM ITS9f primer (final concentration of 0.3 µM), 1 µL of 10 mg/mL bovine serum albumin (final concentration of 1 µg/mL; New England BioLabs, Ipswich, Massachusetts, USA), 0.75 µL of 10 µM barcoded ITS4 primers (final concentration of 0.3 µM), and 1 µL of genomic DNA. Reactions were held at 94 ℃ for 5 minutes to denature DNA, with amplification proceeding for 34 cycles at 95 ℃ for 45 seconds, 50 ℃ for 1 minute, and 72 ℃ for 1 minute and 30 seconds, followed by a final extension for 10 minutes at 72 ℃.

Sequencing libraries were prepared with pooled 16S or ITS amplicons from each sample after purification using Speed Bead Magnetic Carboxylate (GE Healthcare UK Limited, Buckinghamshire, UK) to remove primers. A composite library with equimolar ratios of the purified pooled 16S and ITS amplicons was prepared, and DNA size and quality for sequencing was determined by Qubit and Bioanalyzer (450 ng/ml and average amplicon size of 532 bp, respectively). Custom sequencing primers for 16S and ITS libraries were used as described in ((Caporaso et al., 2012; Looby et al., 2016). The libraries were sequenced by the UCI Genomics High Throughput Sequencing Facility using an Illumina MiSeq platform with paired end reads at 300 bp.

Analysis of 16S and ITS sequencing. The forward reads of amplicon sequences were demultiplexed using QIIME2 version 2018.11 toolkit (Caporaso et al., 2010; Bolyen et al., 2018). Five samples were excluded from the bacterial analysis because of poor sequencing quality, and five samples were removed from the fungal analysis because of duplicated barcodes (Table 1.1 and 1.2). Demultiplexed sequences were denoised using DADA2, with operational taxonomic units, OTUs, picked at 100% identity level (amplicon sequence variants, ASVs) using UCLUST within the QIIME2 pipeline (Caporaso et al., 2010; Edgar, 2010; Callahan et al., 2016;

Bolyen et al., 2018). Resulting OTU tables were rarefied via randomized sampling of sequences without replacement over 300 iterations at a depth of 1,090 and 1,064 sequences per sample for bacteria and fungi, respectively, and using the 'EcolUtils' package in R version 3.6.3 (R., 2018; Salazar, 2020). Taxonomy was assigned to OTUs using bacterial representative sequences and the q2-feature-classifier, classify-sklearn naïve Bayes taxonomy classifier against the Greengenes 13_8 99% OTUs reference sequences (McDonald et al., 2012; Bokulich et al., 2018). Taxonomy for fungal representative sequences were assigned using a dynamic threshold (97 % – 99 % identity to reference), which is based on the most accurate assignment for a given lineage, and determined manually by experts in the field from the UNITE v7.2 database, release date 12-01-2017 (Nilsson et al., 2018; UNITE Community, 2019). Unassigned OTUs at the Kingdom level, or assigned as chloroplasts, mitochondria, and Archaea were excluded from analysis.

Plant community. Species composition and fractional cover was determined in all plots by point intercept during mid-April of 2015, coinciding with late flowering and maximum seed set. Briefly, two 160 × 60 cm^2 polyvinyl chloride frames with 10-cm interval grids were positioned within each plot. A wire was dropped from each grid point, and the first-intercepted species was recorded. The point was recorded as plant litter or bare soil if live plant material was not encountered. The number of interceptions for each plant species was summed within a plot to calculate fractional cover. Fractional cover data of all species observed (32 in total) were used to generate a Bray-Curtis dissimilarity matrix. Fractional cover data for all species was further categorized into five functional groups including; native grasses, non-native grasses, native forb, non-native forb, and native shrub (Kimball et al., 2014; Matulich et al., 2015).

Statistical analysis. To determine the effects of ecosystem, drought, added nitrogen, sample collection date, and all interactions on microbial composition, PERMANOVAs using Bray-Curtis dissimilarity matrices generated with rarefied OTU tables were performed using PRIMER-*e* version 6 (Clarke and Gorley, 2006a; Anderson et al., 2008a). Microbial and plant mixed models included plot treatment (ambient rainfall or nitrogen, drought, added nitrogen, and drought with added nitrogen), ecosystem, and sample collection date as fixed factors. The block factor was nested within ecosystem as a random effect to account for the split-plot design of the experiment. The estimated percentage of variance explained was determined by dividing terms with significant p-values by the sum of the estimates of components of variation given as output from PRIMER-*e*. *Post-hoc* comparisons of PERMANOVAs for drought and collection dates given community dissimilarities were performed using PRIMER-*e*. Multivariate homogeneity of variances for drought and nitrogen treatments by ecosystem were tested in R using the 'betadisper' function of the 'vegan' package, calculating distance to group centroid and accounting for sampling bias (Anderson, 2006; Anderson et al., 2006; R., 2018; Oksanen et al., 2019). To determine which taxa associated with OTU identifiers were key contributors to compositional differences in bacterial and fungal communities, SIMPER tests were conducted in PRIMER-*e*. To visualize factors influencing bacterial, fungal, and plant communities, ordinations of rarefied Bray-Curtis matrices were performed using non-metric multidimensional scaling (NMDS) using the 'vegan' and 'ggplot2' packages in R (Wickham, 2009).

RESULTS

We investigated the response of microbial communities on plant litter to drought and nitrogen addition in a decade long global change experiment carried out in adjacent grassland and CSS ecosystems. From samples taken over three years (beginning near the end of year nine

of the LRGCE) and across 32 treatment plots, 2.48 million bacterial sequences were clustered into 1,197 OTUs (defined at 100% sequence similarity; Table 1.1). The majority of bacterial OTUs were associated with four phyla: Proteobacteria (40.2%), Bacteroidetes (29.6%), Actinobacteria (17.9%), and Firmicutes (2.9%). Correspondingly, 5.69 million fungal sequences clustered into 4,190 OTUs from two main phyla: Ascomycetes (71.4%) and Basidiomycetes (28.2%).

Main effects of ecosystems and time. The composition of both bacterial and fungal communities varied significantly between the grassland and CSS. Ecosystem, including its interactive effects with collection date (encompassing annual and seasonal variation), explained the largest amount of compositional variation, approximately 15% and 10% of the variation for bacterial and fungal communities, respectively (Figure 1.2A and B; Table 1.3; PERMANOVA: $P \leq 0.001$). These compositional differences were apparent at the genus-level. Within bacteria, *Sphingomonas*, *Hymenobacter,* and *Curtobacterium* tended to be relatively more abundant in the grassland, whereas *Janthinobacterium*, *Methylobacterium*, and *Agrobacterium* were relatively more abundant in CSS (Figure 1.3A). Within the fungi, *Alternaria*, and *Vishniacozyma* tended to be relatively more abundant in the grassland, whereas *Cylindroseptoria* and *Coleophoma* were more abundant in CSS (Figure 1.3B). The differences in microbial composition between ecosystems paralleled distinctions in plant community composition, with ecosystem accounting for 38% of the variation in plant composition (Figure 1.4C and D; Table 1.3; $P \leq 0.01$). The grassland was dominated by non-native grasses, whereas CSS was predominantly comprised of native grasses, shrubs and forbs (Figure 1.3C).

Microbial composition also varied temporally over the three sampling years, as expected from previous studies in the grassland at LRGCE (Matulich and Martiny, 2015). After

ecosystem, collection date explained the most variation in both bacterial and fungal community composition, approximately 8% and 2%, respectively ($P \leq 0.001$; Table 1.3; Figure 1.4D). Indeed, microbial communities from sample collection dates coinciding with the peak wet season (e.g., January and March typically have the highest amounts of rainfall) differed in composition from that of collection dates in June and September months coinciding with the peak of the dry season (*post-hoc* pairwise comparisons: $P < 0.01$; Figures 1.1, 1.2, and 1.5).

Main effects of drought and added nitrogen. Drought significantly altered both bacterial and fungal communities (Figure 1.4A and B; Table 1.3; main effects: $P \leq 0.01$), whereas nitrogen only altered the bacterial community. The main effect of drought, including its interactive effects with collection date, explained 5.6% and 3.6% of total variation in bacterial and fungal community composition, respectively (Figure 1.4D). In contrast, added nitrogen accounted for less than 1% the variation in bacterial composition. Main effects of drought also explained 20% of variation in plant community composition across both ecosystems (Figure 1.4C and D). Overall, the main effects of drought were apparent across bacterial, fungal, and plant communities, while the nitrogen treatment only seemed to effect bacterial communities.

Ecosystem-dependent responses to drought and nitrogen addition. In addition to the main effects, the global change treatment altered microbial composition in an ecosystem-dependent manner, in support of our hypothesis. The interaction between the drought treatment and ecosystem explained nearly 5% of the variation in bacterial community composition, similar to the variation explained by the ecosystem-independent effects of drought (main effect plus drought-by-collection date effect = 5.54%; $P = 0.001$; Table 1.3). This interactive effect was apparent at the genus level; for instance, *Curtobacterium* decreased in abundance under drought conditions in the grassland, whereas it increased under drought in CSS (Figure 1.3A). Similarly,

at the OTU level, a number of relatively abundant taxa contributed differentially to compositional shifts under the treatments, responding in opposite directions (positively or negatively) depending on whether they were in the grassland or CSS. Further, some bacterial OTUs were observed exclusively in one ecosystem, where they contributed a large effect to the global change response. For example, OTUs belonging to *Xanthomonadaceae* and *Nesterenkonia* were only detected in grassland plots and increased in response to drought (SIMPER analysis; Table 1.4). These trends illustrate the ways in which bacterial responses to drought can contribute to a significant drought-by-ecosystem interaction.

The response of fungal communities to drought also depended on the ecosystem (drought-by-ecosystem effect: $P = 0.04$; Table 1.3). Like for the bacteria, this ecosystem-dependent response was apparent at the OTU level among the taxa that most contributed to the drought response; for instance, five fungal *Alternaria* OTUs responded to drought in opposite directions for grassland compared to CSS (Table 1.4). Notably, it is unclear if these OTUs also varied significantly among ecosystems (Beta-dispersion analysis: $P < 0.001$; (Warton et al., 2012). However, the fungal response to drought appeared to be less dependent on the ecosystem than the bacterial response; the interactive effect only explained 1% of variation in fungal composition, lower than the variation explained by the ecosystem-independent effects of drought (main effect plus drought-by-collection date effect = 3.58%; Table 1.3).

Like drought, nitrogen addition altered the bacterial community in an ecosystem dependent manner. A nitrogen-by-ecosystem interaction explained a similar amount of bacterial compositional variation (1.1%) as the main nitrogen effect (0.9%). In contrast, the fungal communities did not respond to added nitrogen, either overall or in an ecosystem-dependent manner ($P > 0.05$; Table 1.3).

Finally, parallel to the bacterial and fungal communities, plant community composition shifted in unique ways in the grassland and CSS in response to drought. For example, native and non-native forb cover under drought decreased relative to ambient conditions in grassland plots by approximately 13% and 1%, respectively (Figure 1.3C). Whereas, native forb cover under drought in CSS plots increased by approximately 24% relative to ambient plots. Additionally, ground covered by plant litter in CSS drought plots was nearly three times that of ambient CSS plots, a trend that was not observed in grassland plots (Figure 1.3C). However, drought noticeably increased bare soil cover relative to ambient plots in both ecosystems. Finally, a drought-by-ecosystem interaction explained nearly 16% of variation in plant community composition ($P = 0.001$; Table 1.3), while there was no significant nitrogen-by-ecosystem effect ($P = 0.713$; Table 1.3).

DISCUSSION

After a decade of global change perturbations, microbial communities on decomposing plant litter responded to both drought and nitrogen addition in adjacent ecosystems dominated by different plant communities. When the microbial community responded to the treatments (in all cases but the fungi to nitrogen addition), this response depended in large part on the ecosystem (as indicated by a significant treatment-by-ecosystem interaction), supporting our hypothesis that such responses depend on plant-microbial interactions. Although we cannot separate their contributions here, we suspect that this dependence is due to a combination of the plant community selecting for initially divergent microbial communities, microbial taxa responding differently when situated in different plant communities, and microbial communities indirectly tracking the plant community responses.

After ten years, extreme drought (imposed as a ~50% reduction of annual rainfall) continues to impact both microbial and plant community composition in the LRGCE, as was observed after the first five years of treatment (Potts et al., 2012; Allison et al., 2013; Kimball et al., 2014; Matulich et al., 2015; Kimball et al., 2016; Martiny et al., 2017), and is consistent with drought experiments from other locations (Sheik et al., 2011; Gao et al., 2011; Kinugasa et al., 2012; Schmidt et al., 2018; Griffin-Nolan et al., 2019). Fewer studies consider how the plant communities might alter these drought impacts on microorganisms, but our study contributes to growing evidence suggesting that plant-microbe interactions might be common and play a larger role in microbial drought response than previously thought. For example, the response of bacterial and fungal communities to drought depended on plants, where plant community structure varied by allowing for the "invasion" of grasses (*Imperata cylindrica*) into a longleaf pine (*Pinus palustris*) common garden (Fahey et al., 2020). Evidence further suggests that bulk soil microorganisms influence plant growth under drought conditions. For instance, soil communities selected under drought conditions altered the growth of *Arabidopsis* (Lau and Lennon, 2012), and plant-microbe interactions prior to drought modified the stress response of the grass *Bouteloua gracilis* during drought (Ulrich et al., 2019).

Unexpectedly, we found that the main effect of drought was approximately four times as strong on bacterial composition as that of nitrogen addition, which did not alter fungal or plant composition. Indeed, the differences in plant and microbial community composition in the control and added nitrogen plots appear to be narrowing since the first five years of the treatment. In particular, after five years plant communities in both the grassland and CSS responded to nitrogen addition, albeit not as strongly as drought; nitrogen addition reduced the cover of native grasses and shrubs and increased cover of non-native annual grasses (Kimball et

al., 2014; Kimball et al., 2016). Now after a decade, only the bacterial communities were sensitive to nitrogen and even then, the treatment explained < 1% of the compositional variation, as compared to 2% for both bacteria and fungi at five years (Matulich et al., 2015). However, it is important to note some differences in methodology from earlier studies that preclude more direct comparisons. For instance, previously we characterized fungal diversity using a more conserved gene region, 28S rDNA. The minimal effect of nitrogen fertilization is surprising as it often has large impacts on both plant and soil microbial communities (Elser et al., 2007; Allison and Martiny, 2008; LeBauer and Treseder, 2008; Kinugasa et al., 2012; Legay et al., 2016). We suspect that the attenuated effects of added nitrogen are due to much larger changes in the ambient conditions at the site. Plant composition at the LRGCE is not only shifting in response to the treatments, but also in the ambient plots over time. For instance, native grasses were not detected in any of the grassland plots and have become rare in the CSS plots regardless of their nitrogen status. Indeed, southern California has been subject to a severe long-term drought from 2012 to 2015 (Griffin and Anchukaitis, 2014; Yoon et al., 2015). Hence, the relatively minor effects of added nitrogen may be overshadowed by the larger impacts imposed by prolonged regional drought.

CONCLUSIONS

Global changes such as drought and increased atmospheric nitrogen deposition are likely to alter the composition of both plant and microbial communities (Ciais et al., 2005; Kinugasa et al., 2012; Fuchslueger et al., 2014; Preece et al., 2019; Zhao et al., 2019). A remaining uncertainty, however, is whether microbial responses influenced by changes in plant communities will affect plant community responses, and the predictability of these responses in the long term. Our results add to growing evidence that microbial communities responses to long

term global change such as drought is dependent on biotic factors such as plant communities (Sayer et al., 2017). It is important to note, although we focused here on changes in microbial composition, such changes are often associated with process rates such as decomposition (Strickland et al., 2009; Allison et al., 2013; Cleveland et al., 2014; Martiny et al., 2017; Glassman et al., 2018). Thus, our ability to predict how microbially-driven terrestrial processes will change in the future will require an integrated understanding of both microbial and plant communities (Ostle et al., 2009; Berg et al., 2010; Fischer et al., 2014).

ACKNOWLEDGMENTS

We thank: Alberto Lopez, Ashish Malik, Bahareh Sorouri, Banafshe Khalili, Bin Wang, Chamee Moua, Charlotte Alster, Devon Nelson, Dominique Bao Yang, Joia Kai Capocchi, Jonathan Abebe, Kristin Dolan, Kelly Ramin, Kendra Walters, Mary Xiong, Michaeline Albright, Pauline Uyen Phuong Nguyen, Sydney Glassman, and Zulema Gomez, for field and laboratory help. We also thank the Irvine Ranch Conservancy and UCI Environmental Collaboratory for facilitating research at Loma Ridge. Additionally, the LRGCE is located on the ancestral lands of the Kizh (Jackson and McCawley, 1997) and Acjachemen indigenous peoples.

Table 1.1 Summary of bacterial and fungal community sequencing.

Community and Variable	Pre-rarefied [a]	Post-rarefied [b]								
			Grassland				CSS			
		Total	Ambient Rainfall	Drought	Ambient Nitrogen	Added Nitrogen	Ambient Rainfall	Drought	Ambient Nitrogen	Added Nitrogen
Bacteria										
Total Sequences	2 489 402	209308	52331	47976	50153	50154	54483	54518	55599	53402
Median Sequences per Sample	5558	1090	1090	1090	1090	1090	1090	1090	1090	1090
Total OTUs [c] (100% similarity)	1197	1174	392	365	407	416	642	575	607	631
Relative Abundance of Top 25 OTUs (%)	44	47	57	59	54	57	50	50	51	48
Median OTUs per Sample	32	35	31	38	31	36	37	36	35	37
Mean Shannon Diversity (± s.e.)	-	3.15 (0.04)	3.10 (0.08)	3.03 (0.11)	3.11 (0.09)	3.03 (0.10)	3.26 (0.07)	3.20 (0.07)	3.22 (0.07)	3.25 (0.08)
Fungi										
Total Sequences	5 696 761	214950	56401	51089	58513	48977	51084	56376	51088	56372
Median Sequences per Sample	28031	1064	1064	1064	1064	1064	1064	1064	1064	1064
Total OTUs (100% similarity)	4190	3917	1434	1232	1517	1165	2081	1893	1898	2061
Relative Abundance of Top 25 OTUs (%)	50	50	59	58	57	60	41	48	46	44
Median OTUs per Sample	87	84	76	74	81	70	121	98	101	112
Mean Shannon Diversity (± s.e.)	-	3.92 (0.03)	3.76 (0.06)	3.76 (0.06)	3.83 (0.05)	3.69 (0.06)	4.18 (0.07)	3.97 (0.07)	4.03 (0.08)	4.10 (0.07)

[a] Pre-rarefied samples n = 438.

[b] Post-rarefied samples n = 395; $n_{bacteria}$ = 192; n_{fungi} = 203.

[c] OTUs = Operational Taxonomic Units

Table 1.2 Number of microbial and plant samples included in the study after rarefaction.

	Bacterial[a]		Fungal[b]		Plant	
Survey/Sample collection dates	**Grassland**	**CSS**	**Grassland**	**CSS**	**Grassland**	**CSS**
April 2015	-	-	-	-	16	16
August 2016	10	16	12	14	-	-
December 2016	15	16	15	16	-	-
March 2017	16	16	15	15	-	-
June 2017	9	13	15	13	-	-
September 2017	12	14	16	14	-	-
December 2017	14	9	15	13	-	-
March 2018	16	16	14	16	-	-
Totals	92	100	102	101	16	16

[a] Bacterial samples rarefied to a depth of 1,090 sequences.

[b] Fungal samples rarefied to a depth of 1,064 sequences.

Table 1.3 Results of mixed-model PERMANOVAs of bacterial, fungal, and plant community composition.

Community and Variable	df	SS	MS	Pseudo-F	*P* [a]	% Variance Explained
Bacteria						
Block	6	4.29	0.72	2.66	**0.001**	4.21
Ecosystem	1	5.53	5.53	7.98	**0.001**	11.97
Ecosystem x collection date	6	2.60	0.43	1.61	**0.001**	2.88
Collection date	6	6.98	1.16	4.33	**0.001**	7.54
Drought	1	1.91	1.91	7.11	**0.001**	3.99
Drought x ecosystem	1	1.26	1.26	4.70	**0.001**	4.88
Drought x collection date	6	2.19	0.36	1.36	**0.001**	1.55
Drought x nitrogen	1	0.52	0.52	1.94	**0.001**	1.12
Nitrogen	1	0.64	0.64	2.39	**0.001**	0.89
Nitrogen x ecosystem	1	0.49	0.49	1.83	**0.001**	1.11
Nitrogen x collection date	6	1.67	0.28	1.04	0.346	-
Fungi						
Block	6	1.80	0.30	1.32	**0.029**	1.02
Ecosystem	1	2.12	2.12	7.13	**0.001**	6.63
Ecosystem x collection date	6	2.26	0.38	1.65	**0.002**	3.74
Collection date	6	2.41	0.40	1.76	**0.001**	2.17
Drought	1	0.62	0.62	2.72	**0.003**	1.42
Drought x ecosystem	1	0.37	0.37	1.63	**0.042**	1.04
Drought x collection date	6	1.88	0.31	1.38	**0.015**	2.16
Drought x nitrogen	1	0.36	0.36	1.57	0.067	-
Nitrogen	1	0.22	0.22	0.95	0.472	-
Nitrogen x ecosystem	1	0.22	0.22	0.97	0.404	-
Nitrogen x collection date	6	1.39	0.23	1.01	0.411	-
Plant						
Block	7	1.19	0.17	1.50	**0.046**	3.74
Ecosystem	1	2.70	2.70	15.96	**0.006**	38.19
Drought	1	1.31	1.31	11.56	**0.001**	19.90
Drought x ecosystem	1	0.59	0.59	5.19	**0.001**	15.78
Drought x nitrogen	1	0.06	0.06	0.50	0.838	-
Nitrogen	1	0.08	0.08	0.68	0.654	-
Nitrogen x ecosystem	1	0.07	0.07	0.59	0.713	-

[a] Significant *P*-values ($P \leq 0.05$) are in bold.

Table 1.4 Bacterial and fungal taxa that each contribute to at least 1% of the compositional differences between the treatments by ecosystem.

Class	Nearest Taxonomic ID	OTU ID [a]	Drought Experiment				Nitrogen Experiment			
			Grassland	%	CSS	%	Grassland	%	CSS	%
Bacteria										
Actinobacteria	*Couchioplanes*	bcdf38b56057dd4477e60b11509a7d19	np [c]	na [d]	−	2.25	np	na	−	2.29
	Curtobacterium	1785730613a3462cd513729e4d2105e7	+	5.45	+	2.64	−	5.15	+	2.57
	Frigoribacterium	aaa8848ae680ad429b3ade29166522a1	−	2.31	np	na	+	2.40	np	na
	Frigoribacterium	dc82a6aa614466526b5d0890d10b7d79	−	2.15	+	1.17	+	2.10	+	1.18
	Modestobacter	17cfcd002419bc363616675a2e67a78d	−	0.34	−	1.93	+	0.35	+	1.95
	Modestobacter	9d6d3c667c0ad15eb28de140a3354be1	−	1.44	+	1.66	+	0.46	−	1.64
	Nesterenkonia	95311b9c82dd720c6a03c577a1105598	+	2.98	np	na	+	2.99	np	na
Alphaproteobacteria	*Agrobacterium*	2552d771c4db90b397fb0a8ce11f3f34	−	2.28	np	na	−	2.33	np	na
	Agrobacterium	08395ab1649653023ded96fbdd3c91b4	np	na	−	1.94	np	na	+	2.01
	Methylobacterium	873cdc8f3395a02624e67011ae657a30	np	na	−	4.28	np	na	−	4.35
	Sphingomonas	ff7ac192e37d82ff023e1463105337ce	−	0.34	−	1.64	−	0.35	−	1.65
Bacilli	*Planococcaceae*	579d772237ba3080d015da5a119db073	−	2.67	+	0.52	+	2.39	+	0.53
Betaproteobacteria	*Janthinobacterium*	f76dfc32fd3d75d7e479377c9d12c1c5	−	3.70	+	2.84	+	3.60	+	2.88
	Janthinobacterium	5116c15e3dc5952bc191ed035eb26257	−	2.20	−	4.07	−	2.30	+	4.08
	Janthinobacterium	adaf8937f969e1670c9c0608c3f27a88	np	na	−	2.64	np	na	−	2.48
	Oxalobacteraceae	e54f3924474f1a89a6b227aabedfac19	−	2.13	np	na	+	2.21	np	na
Gammaproteobacteria	*Xanthomonadaceae*	522e67fd09279346ab8fd40defea55fd	+	4.99	np	na	+	4.67	np	na
Cumulative contribution				32.98		27.58		31.30		27.61
Fungi [b]										
Ascomycota	*Alternaria*	e5524ff4bd7d57a6ab6c132e61398402	−	4.36	+	3.00	+	4.40	−	2.99
	Alternaria	bdb1a3b0cc3fff98f06c3af08ef35ea2	−	2.80	+	2.05	+	2.83	−	2.06
	Alternaria	435ec9717b244052bad4ab7904a45f16	−	2.29	+	1.41	+	2.28	−	1.42
	Alternaria	7ca55426aea755263314ec9c4b12084b	−	2.22	+	1.57	+	2.24	−	1.56
	Alternaria	fb7eb06fab9a17ea90b2985feaf2d5bc	−	1.39	+	1.34	+	1.42	+	1.35
	Alternaria	20e03f1ee27c78fd042e9581d27e7f04	+	0.90	+	0.92	+	0.92	−	0.93
	Capnodiales	679f67dc9e1d5ab0f3243e11e77afb9a	−	3.72	+	2.41	−	3.75	−	2.44
	Capnodiales	8f240034c2693d103fc05dac2e10892b	−	2.45	+	1.56	+	2.46	−	1.57
	Capnodiales	52353670278f3f501acd8310a74bcad8	−	1.94	+	1.27	−	1.94	−	1.27
	Cladosporium	0d23c7fb8ec8117dc8553a90e8277d0e	+	2.42	+	3.06	+	2.39	−	3.06
	Cladosporium	866aa412223e98069307981be90829e6	+	1.41	+	1.43	+	1.39	+	1.39
	Cladosporium	910c92d447c4bd4ea4521601b69dfab6	+	1.09	+	1.22	−	1.07	+	1.17
	Coleophoma	fd2020165a09f3df92cddd761d327f0a	+	0.33	−	0.97	+	0.33	−	1.01
	Cylindroseptoria	b54d4643c14f7cdd66f352e4b782ff72	+	0.31	+	2.41	−	0.32	+	1.73
	Leptosphaeria	092401bdba7dffe1659418112f0566be	−	1.21	−	0.47	+	1.22	−	0.45
	Stemphylium	c3397bad751d54869c069fa612505ccd	+	1.29	+	0.84	+	1.28	+	0.83
Basidiomycota	*Vishniacozyma*	0d7bea009c8bd62f8c5b0e9e8fd14907	−	1.01	+	0.43	−	1.00	−	0.43
Cumulative contribution				31.14		26.36		31.24		25.66

[a] Taxa (OTUs, operational taxonomic units) identified using a one-way SIMPER analysis for significant factors identified by the PERMANOVA tests (Table S2). For each factor, the average abundance of the OTUs increases (+) or decreases (−) in the treatment relative to the ambient plots and the OTUs percent contribution (%) to the compositional differences between reduced water or added nitrogen inputs and ambient conditions.

[b] Nitrogen treatments were not significant for fungal communities however trends in fungal taxa are reported.

[c] np, not present in sample.

[d] na, not applicable.

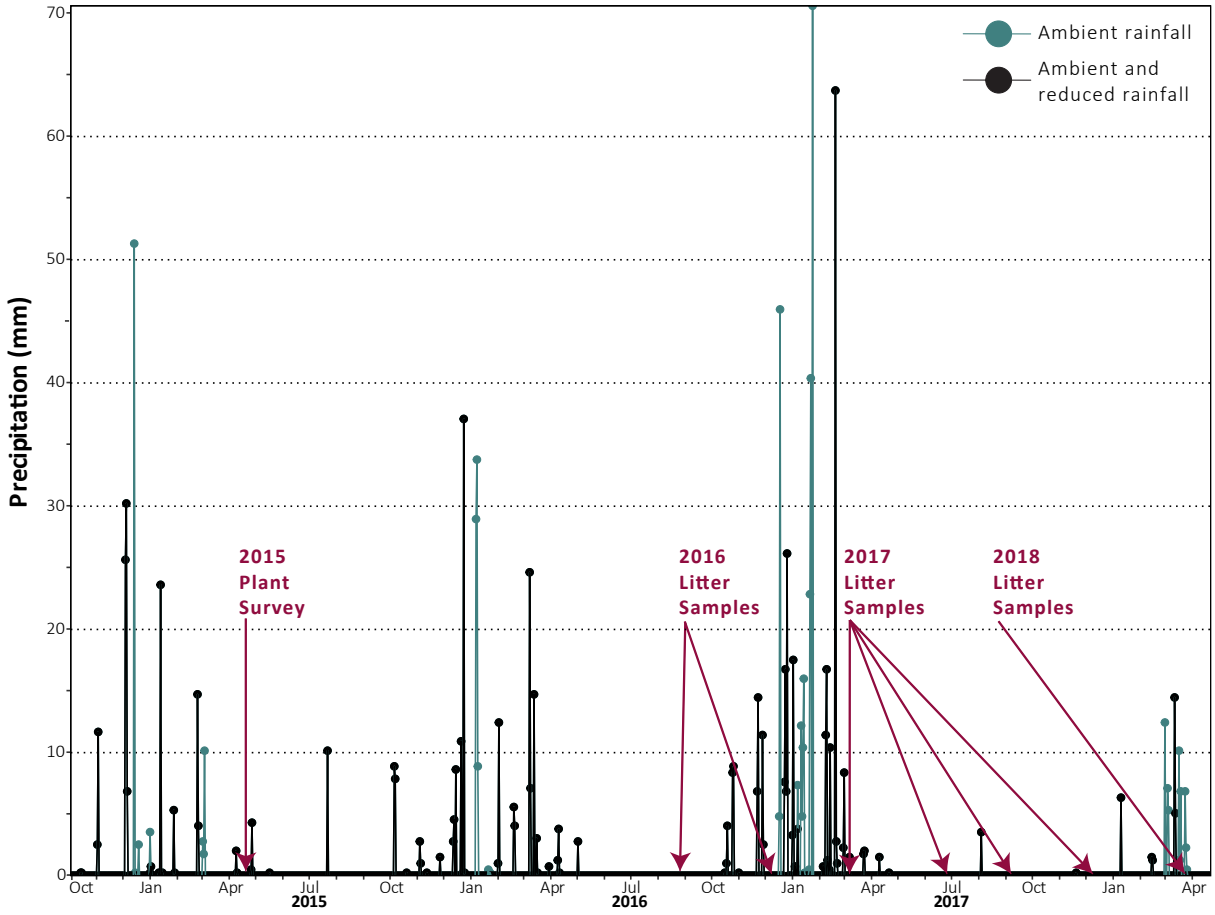

Figure 1.1 Precipitation in the Loma Ridge plots over 4.5 years of a decade long experiment. The green lines denote precipitation that fell only in the ambient (control) plots as it was blocked from the drought (reduced rainfall) plots. Plant survey and litter sample collection dates are marked by red arrows.

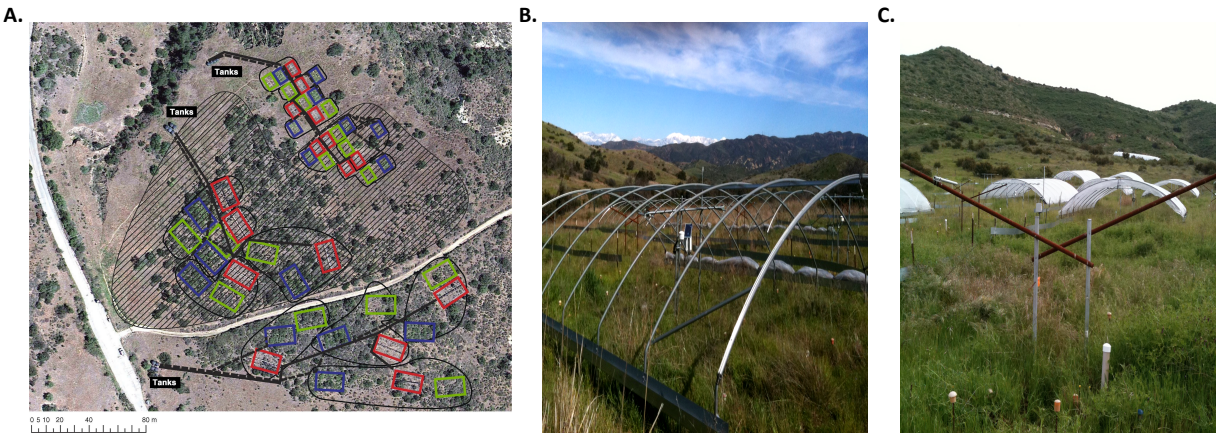

Figure 1.2 Aerial view of Loma Ridge research site and close up of rain shelters atop drought plots. (A) Satellite image of the LRGCE site in the Santa Ana Mountains, within the Irvine Ranch National Landmark in Orange County, California, USA, showing the smaller plots in the annual grassland and the larger, adjacent plots in the CSS, bar scale in bottom left for reference. Crosshatching indicates the area of a February 2007 control burn. Plots are color coded by water treatment, with red indicating reduced rainfall (drought), green for ambient rainfall, and blue for added water (not included in this study). Thick dashed lines indicate water collection pipelines, and solid black lines indicate water distribution pipelines. The black-outlined polygons indicate blocks containing all treatment combinations. Each plot (rectangle) was divided in half lengthwise, and nitrogen treatments (ambient or added) were randomly assigned. (B and C) Image of rain shelters in grassland plots uncovered and covered, respectively.

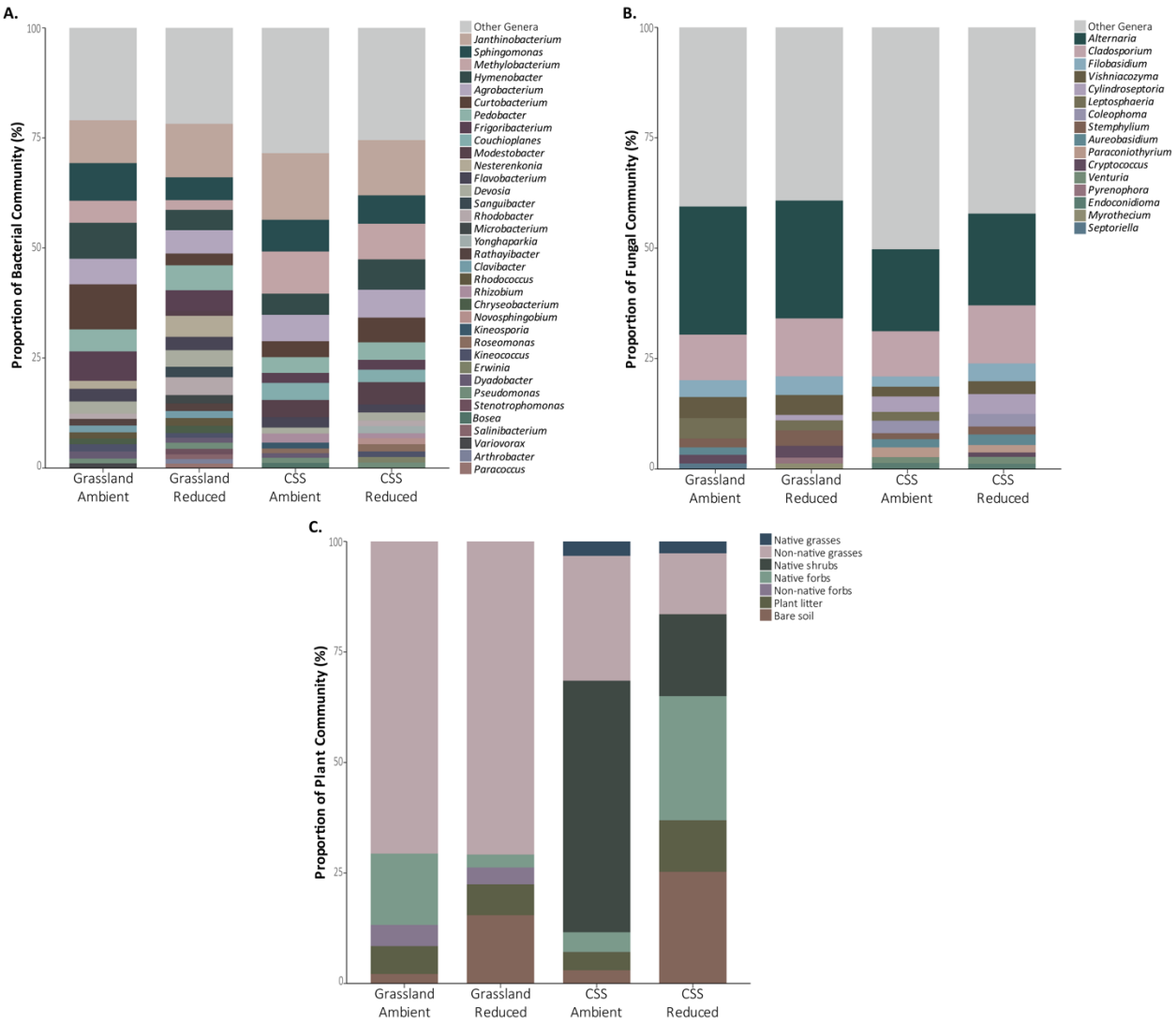

Figure 1.3 Microbial taxonomy and plant function groups differ across ecosystem and drought treatment. Proportional abundances of bacterial (A) and fungal (B) genera and fractional cover of plant functional groups (C) in grassland and CSS ambient rainfall and drought plots (nitrogen not shown because of the minor effects; see text). Microbial taxonomy assigned in QIIME2 using the Greengenes and UNITE databases for bacterial and fungal representative sequences, respectively. All genera under 1% relative abundance or unidentified at the genus level were categorized as "Other Genera".

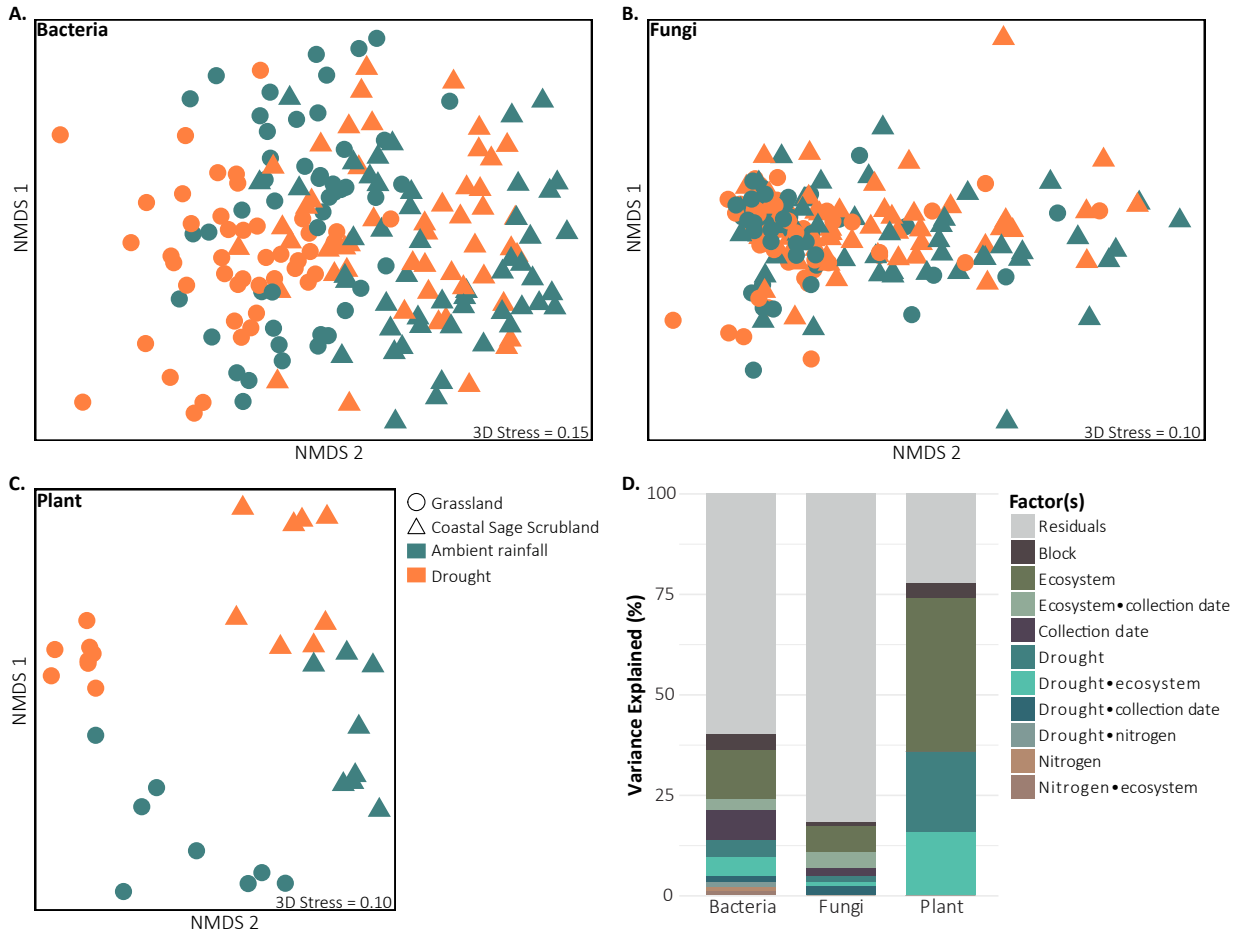

Figure 1.4 Microbial and plant communities vary by drought and ecosystem. Non-metric multidimensional scaling (NMDS) ordination depicting (A) bacterial (B) fungal (C) plant community composition. Symbols are defined in the legend in panel C. Nitrogen treatment and collection date are not plotted here because their effects were not significant or applicable for all communities, see Table 1 and Figure S3. (D) The percentage of variance explained for significant ($P \leq 0.05$; Table 1) factors in a mixed-effects PERMANOVA for bacterial, fungal, and plant community composition.

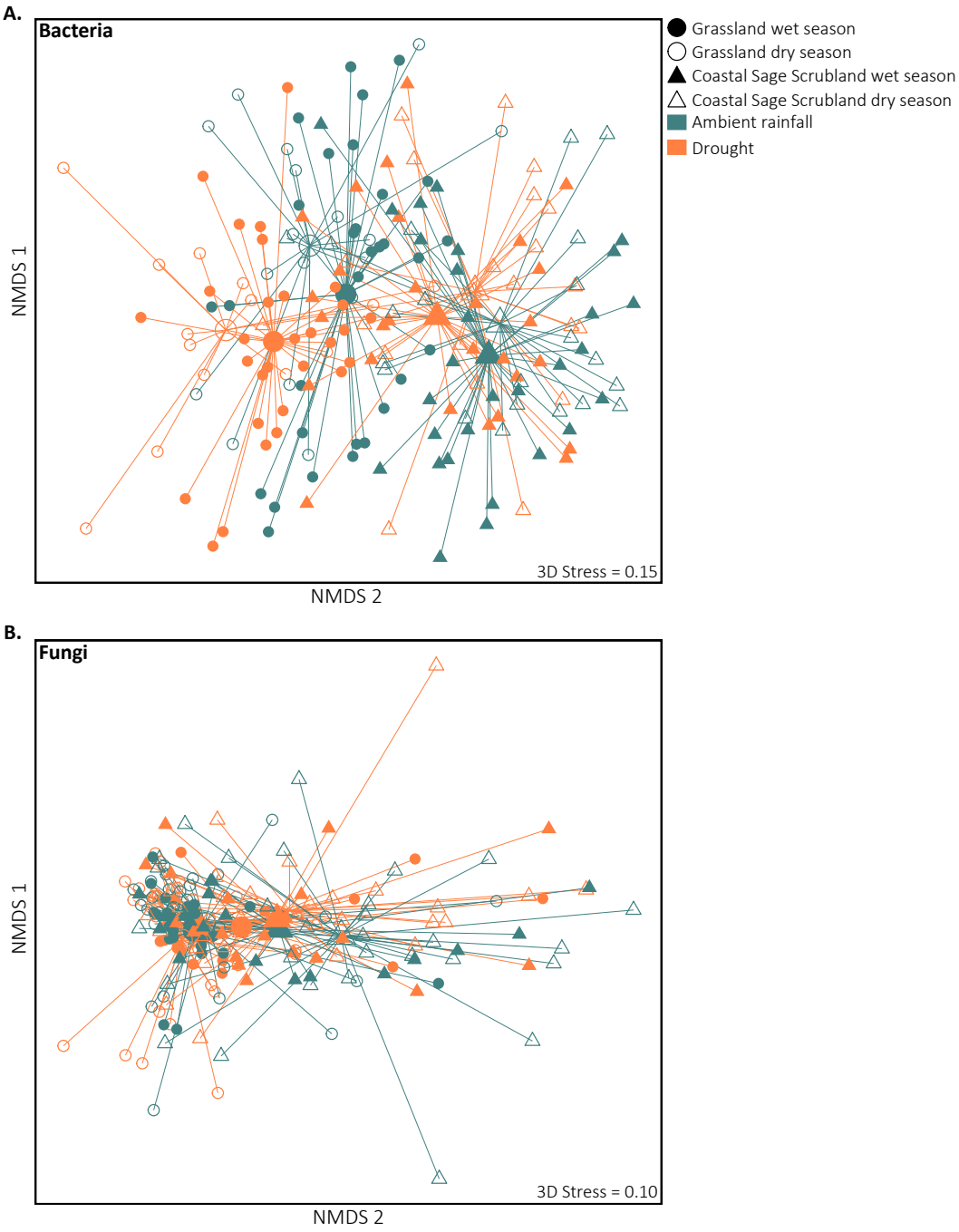

Figure 1.5 Microbial communities are impacted by seasonal variation. Non-metric multidimensional scaling (NMDS) ordination depicting (A) bacterial (B) fungal community composition. Symbols are defined in the legend in panel (A). Centroids are twice the size of sample points. Pairwise PERMANOVAs between microbial communities from sample collection dates coinciding with peak wet (January-March) and dry (June-September) seasons were significantly different $P < 0.01$.

CHAPTER 2

Plasmid-encoded traits vary across environments

ABSTRACT

Plasmids are key mobile genetic elements in bacterial evolution and ecology as they allow for the rapid adaptation of bacteria under selective environmental changes. However, the genetic information associated with plasmids is usually considered separately from information about their environmental origin. To broadly understand what kinds of traits may become mobilized by plasmids in different environments, we analyzed the properties and accessory traits of 9,725 unique plasmid sequences from a publicly available database with known bacterial hosts and isolation sources. Although most plasmid research focuses on resistance traits, such genes made up less than 1% of the total genetic information carried by plasmids. Similar to traits encoded on the bacterial chromosome, plasmid accessory trait composition (including general COG functions, resistance genes, and carbon and nitrogen genes) varied across seven broadly defined environments (human, animal, wastewater, plant, soil, marine, and freshwater). Despite their potential for horizontal gene transfer, plasmid traits were strongly influenced by their host's taxonomic assignment. However, the trait differences across environments of broad COG categories could not be entirely explained by the plasmid host phylum, suggesting that environmental selection may be acting on the plasmid traits themselves. Lastly, some plasmid traits and environments (e.g., resistance genes in human-related environments) were more often associated with mobilizable plasmids than others. Overall, these findings underscore the high diversity of encoded traits by plasmids and provide a baseline for understanding the potential of plasmids to serve as reservoirs of a wide variety of adaptive traits for microbial communities.

IMPORTANCE

Plasmids are well known for their role in the transmission of antibiotic resistance-conferring genes. Beyond human and clinical settings, however, they disseminate many other types of genes, including those that contribute to microbially-driven ecosystem processes. In this study, we identified the distribution of traits genetically encoded by plasmids isolated from seven broadly categorized environments. We find that plasmid trait content was influenced by bacterial host phylum and environment, where on average, half of the plasmids were potentially mobilizable. This analysis greatly enhances our understanding of the types of traits that are mobilized, and the types of host bacteria that may share these plasmid-bound traits. As anthropogenic activities continue to impact ecosystems and the climate, investigating and identifying key mechanisms for how microbial communities will adapt will be imperative for predicting the impacts on ecosystem functioning.

INTRODUCTION

The acquisition of new traits from mobile genetic elements such as plasmids is broadly thought to be important to bacterial diversification and adaptation (Wollman and Jacob, 1956; Kirchberger et al., 2020). Plasmids carry genes that encode a diversity of traits involved in plasmid-specific functions as well as those related to the physiology of their host (Garcillán-Barcia et al., 2009; Rankin et al., 2011; Pinto et al., 2012; Stasiak et al., 2014). However, analyses of bacterial traits and their corresponding ecological roles typically focus on chromosomal genetic content (Smalla et al., 2015). Furthermore, in non-clinical microbial genomics studies, plasmids are often not distinguished from chromosomal sequences or inadvertently removed, potentially obscuring our understanding of trait variation in environmental communities (Brito, 2021). Indeed, understanding what kinds of traits are carried

by plasmids and why, remains an open question in plasmid ecology (Brito, 2021; Brockhurst and Harrison, 2021). Some evidence suggests ecology (or the local biotic and abiotic environment), rather than geographic isolation and host phylogeny, drives plasmid-mediated gene exchange in bacteria (Smillie et al., 2011; Stecher et al., 2012; Bruto et al., 2017). To begin to understand the extent to which some traits may be limited by ecological opportunity and occupancy of shared habitats, a first step is to characterize trait variation on plasmids across different environments. Many studies have focused on the intrinsic properties of plasmids, including their structure (linear or circular forms), size, copy number, mechanism of replication and segregation, GC content, coding density, host-range, and mobility (Smillie et al., 2010; Shintani et al., 2015; Hall et al., 2022), but this information is usually considered separately from their environmental origin. Indeed, much of our understanding of the diversity and ecological significance of plasmid properties is derived from a limited number of bacterial phyla and environments, particularly those of *Proteobacteria* and *Firmicutes* that were isolated from human and other host-associated environments. Finally, these studies generally do not address the full diversity of accessory traits (Hall et al., 2017), or traits encoded by plasmids that are generally associated with bacterial chromosomes and thus, may provide their host with a selective advantage.

Perhaps the best-studied plasmid accessory traits are those for antibiotic resistance and virulence (M. et al., 1999; Li et al., 2015; Hu et al., 2016; San Millan, 2018). For instance, plasmids carrying and transferring genes encoding extended spectrum beta-lactamases are responsible for multidrug resistance in pathogenic bacterial hosts (Jacoby and Sutton, 1991; Philippon et al., 1994; Nordmann et al., 2011; Carattoli, 2013). Plasmid virulence genes, such as those encoding fimbriae (important for bacterial attachment to the human reticuloendothelial system), underlie the ability of bacterial pathogens to establish infections (Rotger and Casadesús,

1999; Farshad et al., 2012). More broadly, resistance genes – involved in resistance to heavy metals, biocides, and antibiotics – are commonly reported from plasmids in human-impacted environments. For example, plasmids from wastewater treatment plants are considered reservoirs of resistance traits (Stalder et al., 2019). Similarly, plasmids from environments contaminated with petroleum and other pollutants encode resistance to antibiotics, heavy metals, and herbicides, as well as pathways to degrade xenobiotics and protect against ultraviolet radiation and exogenous DNA (Sayler et al., 1990; Fulthorpe and Wyndham, 1991; Collard et al., 1994; Nakatsu et al., 1995; Sen et al., 2011; Heuer and Smalla, 2012; Romaniuk et al., 2018; Moretto et al., 2019; Dunivin et al., 2019).

Even in less-heavily human-impacted environments, plasmid accessory traits reveal a connection between plasmid genetic content and their ecological roles. The bovine-rumen plasmidome (the overall plasmid population of microbial ecosystems using culture-independent methods) includes genes enriched for amino-acid, cell-wall and capsule, vitamin, and protein metabolism functions, suggesting their importance in conferring nutritional advantages to their bacterial hosts (Brown Kav et al., 2012). In rhizosphere bacteria, plasmid genes for nitrogen fixation aid in establishing symbiotic relationships between plant and host-bacteria (Long, 1996; Davison, 1999; Masson-Boivin et al., 2009). Plasmids from aquatic environments are generally less studied (Davison, 1999), but those from marine *Lentimonas sp*. encode putative carbohydrate active enzymes (CAZymes), including fucoidanases, glycoside hydrolases, sulfatases and carbohydrate esterases, are important for degrading recalcitrant polysaccharides (Sichert et al., 2020). Yet despite these individual examples, it remains unclear how plasmid traits contribute to the overall genetic diversity of bacterial communities across environments.

To begin to address this knowledge gap, we analyzed a publicly available database having over 23,000 plasmid sequences (Galata et al., 2019). The PLSDB, is a resource containing plasmid records collected from the NCBI nucleotide database and allows for gene searches, comparing plasmids and containment analysis. However, we note that such resources are depended on cultured strains and thus, represent a subsample of the diversity of bacteria and their plasmids across environments. We also note that some similar plasmids may be sampled multiple times from particular species or environments, as part of a targeted research project into those taxa and environments, or as a side effect of the locations of research investigating these plasmids. Nevertheless, analysis of databases such as this can help further our understanding of the genetic diversity that is encoded by plasmids and the insight into how plasmid accessory genes vary by environment. Given the potential for plasmids to influence the ecological roles of bacterial communities, we first asked: do plasmid properties and accessory traits vary by environment? We characterized plasmid accessory traits generally, assigning clusters of orthologous genes (COG functions) to all gene calls, but also focused on three specific gene functions of interest: resistance genes, carbon degradation genes (CAZymes), and inorganic/organic nitrogen processing genes. While the latter two functions are not often considered as plasmid-associated, they are central to biogeochemical processes and ecosystem functioning. We hypothesized that plasmid accessory traits are subject to similar selection pressures experienced by the host microbiomes for these traits (Berlemont and Martiny, 2013; Nelson et al., 2016) and would therefore vary by environment.

Secondly, we asked: to what extent does host taxonomy influence plasmid accessory traits across environments? Because the taxonomic composition of bacterial communities varies dramatically across environments, it is difficult to fully separate the influence of the environment

versus bacterial host taxonomy on plasmid traits. However, to start to disentangle factors influencing plasmid accessory trait content, we compared the effect of bacterial environment versus plasmid environment (host bacterial phylum) on plasmid trait composition. Bacterial host taxonomy is represented by phylogenetic relationships between bacterial hosts, which may also influence the relatedness of plasmids resulting in non-independence between samples (Dewar et al., 2021). However, previous analysis of over 10,000 reference plasmids has revealed that more than 60% of the plasmids group by host ranges beyond the species barrier (Redondo-Salvo et al., 2020), here we assume the majority of plasmids follow independent evolutionary trajectories from bacterial species. Simply put, plasmids, like their hosts, form coherent genomic groups similar to molecular species. Whether or not plasmid phylogeny tracks with host species phylogeny is not addressed in this paper. Instead, we focus on addressing the impact host phylum might have on plasmid accessory trait content. Thus, we hypothesized that plasmid accessory genes, would be more strongly influenced by the environment a bacterial host was isolated from which may be subject to various environmental selection pressures (Rizzo et al., 2013; Berendonk et al., 2015), rather than host taxonomy. Recent analysis of some plasmid properties have revealed host taxonomic influence on these traits (Redondo-Salvo et al., 2020), whereas the environment may influence plasmid accessory trait content (Perez et al., 2020). Alternatively, Redondo-Salvo et al., 2020 examined the homologous protein clusters (HPC) found on plasmids at different bacterial-host taxonomic levels and found the fraction of HPC decreased with phylogenetic distance, especially above the order rank (Redondo-Salvo et al., 2020). This analysis suggested that host similarity acts as a constraint for the propagation of plasmid genetic information. We also note that plasmid mobility may also be influenced by co-evolutionary dynamics with their host which may have some impact on accessory gene content (Harrison and

Brockhurst, 2012; Hall et al., 2021). Thus, it is likely that both environmental and host phylogeny act together to shape the plasmid accessory content, it is our aim to begin to quantify the extent of these forces on plasmid traits.

Finally, we asked: are certain accessory traits and/or environments more often associated with mobilizable plasmids? We anticipated that some accessory traits would be more often associated with mobilizable plasmids (those carrying the genes necessary to hitchhike with self-transmissible/conjugative plasmids during transfer to other bacteria). For instance, some environments are known reservoirs of antibiotic resistance traits such as humans and wastewater (Rahube et al., 2014; Craig et al., 2022), and in such environments, the capture of resistance-conferring genes by plasmids may be facilitated by interactions with other mobile genetic elements and host bacteria (Sentchilo et al., 2013; Ghaly et al., 2017).

MATERIALS AND METHODS

Retrieval of plasmids from different environments. Plasmid sequences and metadata were retrieved from the curated plasmid database PLSDB v.2020_06_29 containing 23,227 on 2020-10-14 (Galata et al., 2019). Plasmid sequences from seven environment types (human, animal/non-human, wastewater, plant, soil, marine, and freshwater) were obtained by using the PLSDB provided metadata: *IsolationSource_BIOSAMPLE*, *Host_BIOSAMPLE*, and *SampleType_BIOSAMPLE*. Plasmids from human environments were identified by search terms: human, OR *Homo sapiens*, OR child, OR patient. Plasmids from animal and plant environments were identified by both common and/or scientific names (e.g., mouse, OR mice, OR *Mus musculus*). Plasmids from wastewater were identified by the search terms: wastewater, OR sewage, OR sludge. While plasmids from soil, marine, and freshwater environments were identified by the terms: soil(s), OR mud, OR permafrost, and marine, OR ocean, OR sea, OR

seawater, OR beach, and freshwater, OR lake, OR river, OR creek, OR stream, respectively. Plasmids from rhizosphere, OR root, OR root nodule were assigned as plant environment, while sequences identified as 'rhizosphere soil' were assigned to the soil environment. Duplicate plasmid sequences were removed using the unique plasmid record identifier: *UID_NUCCORE*. If *BIOSAMPLE* categories resulted in the same accessions being binned into different environments, the plasmid environment assignment was determined first by sample type, then by host (e.g., human or animal), and lastly by isolation source. Plasmid sequences were retrieved from PLSDB database files using the *UID_NUCCORE* identifier and the *blastcmd* feature of BLAST+ v2.10.0 (Camacho et al., 2009).

Plasmid properties and accessory trait identification. To identify plasmid sizes (nucleotide length), the *Length_NUCCORE* of PLSDB metadata was used[3]. Depending on the sequence format of the databases used for trait identification, different search tools were employed. Plasmid sequences were searched for MOB family relaxases, which are essential for conjugative DNA processing (Garcillán-Barcia et al., 2009) using MobScan[4] (Garcillán-Barcia et al., 2020), after assigning gene calls using Prodigal v2.6.3 (Hyatt et al., 2010). MOB hits with per-domain thresholds (i-Evalues) ≤ 1e-5 and > 60 % query coverages were included in the analysis.

To identify pathway and functional systems encompassing a diversity of traits, the Clusters of Orthologous Genes database, release 2020 (Galperin et al., 2021) was searched using plasmid gene calls (excluding partial gene calls) and DIAMOND v0.9.14 in sensitive mode (Buchfink et al., 2015). All COG hits with E-values ≤ 0 were included in the analysis, and in

[3] https://ccb-microbe.cs.uni-saarland.de/plsdb/plasmids/download/
[4] https://castillo.dicom.unican.es/mobscan_about/

cases where multiple domains for a given gene call resulted in more than one COG function and category assignment, only annotations for the first domain hit were included in the downstream analyses. For identifying traits involved in biogeochemical processes (carbon and nitrogen cycling pathways), heavy metal and resistance determinants, similarity searches of plasmid gene calls against the following databases were used: the standalone version of dbCAN2, release 2019-07-31 (Zhang et al., 2018), the NCycDB – with curated nitrogen (N) gene family sequences (encompassing seven N-cycling pathways) at 100 % sequence identity, release 2019 (Tu et al., 2019), and MEGARes version 2.0.0, a database for classification of heavy metals, biocides, antimicrobials, and antibiotic resistance determinants (Doster et al., 2020). For carbohydrate utilization trait analyses, CAZymes hits in the dbCAN2 database were included if two or more of the three search tools (HMMER, DIAMOND, Hotpep) matched in their identification of the same CAZyme family. For instances where a single gene call returned multiple matches to CAZyme families, only the annotations from the first domain hit were included in downstream analyses. For N-gene family hits, BLASTp searches of plasmid gene calls having E-values $\leq 10^{-5}$ and > 50 % query coverages were included in the analyses. Since N-gene families encode for a single N-cycling pathway in the NCycDB, these terms were used interchangeably throughout the paper. For resistance determinant identification, BLASTn searches of plasmid gene calls having E-values $\leq 10^{-5}$, and > 85 % query coverage per subject and high-scoring pairs, were included in the analyses.

To determine the extent to which accessory traits are mobilized via plasmids, we first designated plasmids into two mobility categories (mobilizable and nontransmissible) based on previous assessments of plasmid mobility (Smillie et al., 2010). However, here we do not distinguish self- transmissible/conjugative plasmids from mobilizable. Since known plasmid

DNA relaxases (MOB genes) with or without genes encoding type IV coupling proteins are presumed mobilizable, while conjugative plasmids require the major components of a type IV secretion system (T4SS) in addition to MOB genes to be self-transmissible, we broadly define the potential of plasmids to move to other bacteria based on the presence or absence of known MOB genes and focus on the associated accessory trait distributions by environment. We refer the reader to previous studies for an in depth review on the quantification and diversity of T4SS and plasmid mobility (Alvarez-Martinez and Christie, 2009; Smillie et al., 2010).

Plasmid accessory trait and clustering analysis. To standardize for uneven plasmid sequences across environments, trait counts were first converted into proportional abundances within an environment after removal of rare traits (traits counts < 6 across all environments). Trait abundance across environments were then normalized using Z-scores in R v3.6.3 (R Core Team (2020), 2020). This procedure served to weigh each trait similarly, rather than proportional to its abundance. To compare the similarity of traits across environments, we then calculated the Euclidean distance of the standardized and normalized trait count data using the *vegdist* function of the 'vegan' package in R (Oksanen et al., 2019). To determine trait clustering by environment and trait category, agglomerative hierarchical clustering of the distance matrices was performed using the *hclust* function (clustering method = "average" for UPGMA) of the 'stats' package in R. To visualize trait clustering results, heatmaps using trait distance matrices were passed to *heatmap.2* function of the 'gplots'[5] package in R. To determine the extent to which total plasmid trait diversity is characterized across plasmid environments, cumulative COG richness was assessed for unique COG function accessions grouped into broader category designations, which

[5] https://github.com/talgalili/gplots

were subsampled across each environment using the *rarecurve* function (step size = 1,000) of the 'vegan' package in R.

Statistical analysis. To assess the differences in how plasmid size and coding density may vary by environment and host taxonomy (hereafter "host" refers to the bacteria from which plasmids were isolated and is distinct from host-associated environments, e.g., human and animal), we used Kruskal-Wallis tests on $\log_{10}$ transformed nucleotide length (bp) of plasmid sequences grouped by environment or host phyla and using the 'stats' package of R. Post-hoc Wilcoxon tests (with Bonferroni-adjusted *P*-values for multiple comparisons) of plasmid sizes by environment, phyla, and environment given a particular phylum were performed using the *pairwise.wilcox.test* function of the 'stats' package in R. Plasmid coding density, was calculated as the ratio of gene calls per nucleotide length in kilobases (Land et al., 2014). To assess the relationship between plasmid size and coding density, a Kendall rank correlation coefficient (τ) were obtained using *cor.test* function of the 'stats' package in R. To disentangle the influence environment and/or host-taxonomy had a plasmid trait composition, permutational multivariate analysis of variance PERMANOVA (permutations $n = 999$ under a reduced model) on Euclidean distance matrices of standardized and normalized trait counts as previously mentioned, were performed in PRIMER-e v6 (Anderson, 2001; Clarke and Gorley, 2006a; Anderson et al., 2008a). We used a plasmid accessory-traits in fixed-effects model which included environment and host-phylum, with no interaction term, partial sums of squares, and fixed effects sum to zero to test for the effect (significance and variance explained) of host taxonomy versus environment on the composition of each trait type. Since plasmids from some host taxa were represented across few environments (e.g., plasmids of *Chlamydiae* were present in human and animal environments only), we tested the influence host-phylum has on accessory trait content for

plasmid host taxa with at least 50 representatives in the dataset. The percentage of estimated variance explained for significant factors were determined by dividing terms by the sum of the estimates of components of variation and multiplying by one hundred. To determine that the assumptions of PERMANOVA tests were met, PERMDISP, distance-based test for homogeneity of multivariate dispersion, in PRIMER-e were performed (permutations $n = 999$) measuring deviations from group centroids (Clarke and Gorley, 2006b; Anderson et al., 2008a). To determine whether the proportions of traits were different among environments, Log-Likelihood-Ratio (G-test) test of independence using the *GTest* function (with no correction) of the 'DescTools'[6] v0.99.44 package in R were performed. G-tests were performed on contingency tables of non-standardize trait counts with rare traits (traits counts < 6 across all environments) removed. Stacked bar plots of standardized MOB gene counts by environment and linear plots of plasmid coding densities by environment were constructed using 'ggplot2'[7].

Sensitivity analyses. To test the sensitivity of our results (that accessory gene content differed by environment) to sampling biases of the plasmid database, we repeated our main analyses on subsets of the data. To do this, we calculated the taxonomic distribution of bacterial hosts in the dataset (Figure 2.1) using the r package 'ggsankey'[8]. First, we investigated patterns within the most abundant phyla, as explained above. Then, we examined trends within *E. coli*, the most abundant species represented. Finally, we removed the three most abundant genera: *Acinetobacter*, *Escherichia*, and *Klebsiella*, from *Proteobacteria* and *Staphylococcus* from *Firmicutes* and tested that the results held within these phyla.

RESULTS

[6] https://andrisignorell.github.io/DescTools/
[7] https://ggplot2.tidyverse.org
[8] https://github.com/davidsjoberg/ggsankey

After filtering the PLSDB by our criteria, we analyzed 9,725 unique plasmid sequences from human, animal, wastewater, plant, soil, marine, and freshwater environments. As expected, the environmental and taxonomic representation of the plasmid sequences was highly skewed (Figure 2.1 – 2.2). The majority (52.6 %) of plasmid sequences were from human environments, whereas 11% and 2% of plasmids were from plant and marine environments, respectively (Table 2.1). Moreover, the bulk of the plasmid sequences were associated with three bacterial host phyla, namely *Proteobacteria* (n = 7,319), *Firmicutes* (n = 1,565), and *Actinobacteria* (n = 304).

Plasmid properties vary by environment. Plasmid properties including size, coding density, and mobility differed by environment. Overall, plasmid sizes varied by environment (Kruskal-Wallis: H (6) = 762.23, $P < 0.001$; Figure 2.3), and particularly between some environments such as human-plant, and wastewater-marine (Wilcoxon with Bonferroni adjustments: $P < 0.01$). Plasmid size also varied by phylum (H (12) = 521.9, $P < 0.001$; Figure 2.4A); where the mean size of *Proteobacteria* plasmids were significantly larger (146 kb) than *Bacteroidetes* (73.1 kb), *Firmicutes* (68 kb), and *Chlamydiae* (7.5 kb) plasmids (Wilcoxon tests: $P < 0.001$). However, size differences between environments were not entirely attributable to phylum differences as plasmid sizes differed across environments within phyla. For instance, plasmid sizes within *Proteobacteria* (the most abundant host phylum in the dataset) also differed by environment (H (6) = 717.99, $P < 0.001$; Figure 2.4B).

Across all plasmids, we identified over 1 million gene calls with an average coding density of 1.07 genes per kb. Furthermore, the mean coding density varied significantly by environment (H (6) = 360.74, $P < 0.001$; Figure 2.5) with slightly higher coding densities in humans, animals, and wastewater, compared to other environments. Notably, smaller sized

plasmids tended to have greater coding densities than larger ones (Spearman's $\rho = -0.11$; $P < 0.001$), contrary to previous findings (Smillie et al., 2010).

The most abundant genes encoding for plasmid mobility (specifically, the MOB family of relaxases) were MOBP, MOBB, and MOBF (Figure 2.6). Moreover, the percentage of mobilizable plasmids (those having at least one MOB identified) ranged from 41 – 60 % across environments, with the lowest percentage found in the marine and the highest in wastewater environments (Table 2.1). MOB composition also varied by host-phylum (PERMANOVA, $P = 0.001$), but not by environment (Figure 2.6; Table 2.2). Indeed, ~ 44 % of the variation in MOB composition was explained by bacterial host phylum, however, we note that this could be due to differences in average composition and/or within-phylum variance (PERMDISP: $P = 0.004$; ref. 70).

Plasmid accessory traits vary by environment. Overall, ~ 56 % of the plasmid gene calls were assigned to known COG functions, excluding gene calls assigned to general or hypothetical functions (Table 2.1). Most gene calls were classified into 25 broader COG categories, with thousands of functions observed in plasmids from all environments, with soil plasmids having the greatest (> 3500) functional diversity (Figure 2.8). However, despite the large number of COG functions identified, the diversity of COG functions remained undersampled for some environments. In particular, the identification of new COG functions continued to increase with each plasmid gene observed, with steeper rates in aquatic environments than in plant and human environments (Figure 2.8). Noticeably, ~ 38 % of gene calls did not match any known COG functions, suggesting that plasmid gene content captures novel gene products and functions (Table 2.1).

To investigate how these functions varied by environment, we first considered broad COG categories (assigned A through X) of the plasmid genes. Indeed, the prevalence of COG categories differed significantly by environment (G-test: G (138) = 140,898, $P < 0.001$; Figure 2.9). Most COG categories were detected on plasmids across all environments and revealed distinct trends in their relative abundance. For example, COG categories of replication (L), defense mechanisms (V), recombination and repair (L), and mobilome (X), were more prevalent in plasmids from wastewater, animal, and human environments. In contrast, COG categories for carbohydrate/amino acid/nucleotide transport and metabolism (G/E/F) and transcription (K) were more prevalent in plasmids from plant, marine, freshwater, and soil environments. Plasmids from plant, marine, freshwater, and soil had a higher prevalence of genes assigned to general and hypothetical functions relative to those in wastewater, animal, and human environments.

More than 4,800 genes encoding resistance were identified, accounting for 0.72 % of the plasmid accessory traits across all environments (Table 2.1). Of these genes, the majority conferred resistance to antibiotics (58 %), followed by resistance to heavy metals and biocides (38 %) and antimicrobials (4 %). As with overall COG categories, the composition of plasmid resistance genes also differed significantly by environment (G-test: G (138) = 1058.9, $P < 0.001$; Figure 2.10A). Not surprisingly, resistance genes associated with commonly prescribed classes of antibiotics (trimethoprim, fluoroquinolones, tetracyclines, and phenicol) were relatively more prevalent in plasmids from animal and human environments, and some classes (beta-lactams and aminoglycosides) of resistance were found throughout all environments. In contrast, plasmid resistance genes for heavy metals and biocides (metal, nickel, arsenic, and copper) were generally more prevalent in plasmids from plant, soil, freshwater, and marine environments. One

exception to this trend, however, was that resistance genes for lead were relatively more abundant in plasmids from human, animal, and wastewater environments.

Genes encoding CAZymes represented 2.6 % of plasmid accessory traits identified. Glycoside hydrolases (GHs) and glycosyltransferases (GTs) were the most common, accounting for 53 % and 20 % of the CAZymes identified (n = 133), respectively. The least prevalent plasmid CAZymes (< 9% across environments) were those of polysaccharide oxidases (AA), carbohydrate binding modules (CBM) and pectate lyases (PL). As with the other accessory traits, all types of CAZymes (not just GHs) were present in all environments but varied distinctly in prevalence by environment (G (786) = 10335, $P < 0.001$; Figure 2.10B). For instance, GT2, GH1, and GH3, were present across most plasmid environments, with GT2 more prevalent in plasmids from freshwater, soil, and marine environments, while GH1 and GH3 was more prevalent in freshwater environments. Some CAZyme types were present almost entirely in one plasmid environment. For example, GH91, AA7, GT23, and PL3_2 were highly prevalent in plasmids from plant environments, while GH5 and GH32 were highly prevalent in plasmids from human environments. Although, some GH5 subfamilies, such as GH5_48 were highly prevalent in plasmids from marine environments.

Nitrogen gene families represented 0.59 % of plasmid accessory traits identified (Table 2.1). The majority of plasmid N-gene families were assigned to two main N-cycling pathways: organic degradation and synthesis (~ 59 %) and denitrification (21.5 %; Figure 2.10C). In general, N-gene families and their corresponding N-cycling pathways were present across most plasmid environments, with distinct trends in prevalence by environment (G (192) = 1,049, $P <$ 0.001). For instance, *napC* and *nirK*, encoding denitrification functions were present across most environments, and highly prevalent in freshwater and marine environments, respectively. In

contrast, a few N-gene families were identified in plasmids from only one environment; for instance, the nitrogenase, *nifH* and *gdh_K00260*, glutamate dehydrogenase were identified only in plant environments. Of note, some plasmid N-gene families encoding for nitrification and anaerobic ammonium oxidation were rarely identified or not detected in this dataset.

Plasmid accessory traits vary by host phyla. Given that bacterial community composition varies tremendously across environments, the distribution of plasmid accessory traits could be largely driven by changes in host composition rather than direct selection on the plasmid traits themselves. Indeed, plasmid accessory trait composition varied significantly by host phylum for all trait types; however, the percent variance explained by host taxonomy ranged widely depending on the trait type (Table 2.2). In particular, host phylum explained 40 % of compositional variation in COG category traits, whereas it explained much less for resistance (13 %), carbon (1 %) and nitrogen cycling (10 %) trait composition. When controlling for host phylum, 9 % of the variance in COG composition was explained by environment (PERMANOVA: $P = 0.04$), and similar trends by environment were apparent for the COG categories amongst the three most abundant phyla (Figure 2.11). However, the trait composition of resistance genes, CAZyme, and N-gene families did not vary significantly across environments, although we suspect this may be due to poor sampling of particular traits in some phyla (Figures 2.12 – 2.14).

Mobility potential of plasmid accessory traits. We next asked if some traits or environments were more often associated with potentially mobilizable plasmids and identified which accessory traits were on plasmids having at least one gene encoding a MOB relaxase (required for both conjugative and mobilizable plasmids). Approximately half of all accessory traits were identified in plasmids having at least one MOB gene (Table 2.3). The highest average

percentage of potentially mobilizable genes were identified in resistance (64.4 %) traits compared to COG (46.4 %), carbon (45.5 %) and nitrogen (44.9 %) cycling, respectively. Further, plasmids from human, animal, and wastewater environments had relatively higher average percentages (e.g., 58.8 % of plasmid COGs from human vs. 31.6 % in soil, G (6) = 26,449, $P < 0.001$) of mobilizable traits compared to the other environments, except for N-cycling traits. In that case, plasmids from plant environments had the highest percentage (53.6 %) of mobilizable N-gene families.

The distribution of potentially mobilizable traits also varied widely and significantly among specific traits within trait types (Tables 2.4 – 2.6). For instance, over 65 % of plasmid genes encoding resistance to beta-lactams, naphthoquinone, and sulfonamides drug classes were associated with mobilizable plasmids compared to those encoding resistance to rifampin and tetracycline classes with < 50 %, respectively (G (27) = 124.3, $P < 0.001$; Table 2.4). Similarly, GH23 and PL3_1 CAZymes (80 % and 90 %, respectively) were more often associated with mobilizable plasmids compared to GH1 or PL9_2 (39 % and 5 % mobilizable, respectively; G (281) = 4593.7, $P < 0.001$, Table 2.5). Amongst the N-gene families - nitrogen fixing genes: *nifD, nifH, nifK, and nifW* were associated with the highest percentage (93%, 89%, 85%, and 100%, respectively) of mobilizable plasmids compared to genes encoding denitrification pathways, such as *napA* and *napC* with approximately 33% and 44% associated with mobilizable plasmids, respectively (G (39) = 296.6, $P < 0.001$, Table 2.6).

Trait variation is not entirely explained by taxonomic biases. The distribution of host taxonomy across the PLSDB dataset is highly biased (Figure 2.1). A handful of genera are heavily overrepresented, with the top three genera of *Escherichia, Klebsiella,* and *Acinetobacter* accounting for 41% of the plasmids we analyzed. The biases are similarly problematic at broader

taxonomic categories; for instance, the family *Enterobacteriaceae* makes up 44 % of the plasmids and the phylum *Proteobacteria*, 75 %.

The above results demonstrate that accessory gene content varied across environments even within the most abundant phyla in the dataset. To further test the sensitivity of the results, we also tested patterns within the most abundant host species (*E. coli*, 19 % of all plasmids). The bulk of *E. coli* plasmids were from animal and human environments (36 % and 61 %, respectively) and no representatives were detected from soil. However, COG composition varied significantly between human- and non-human animal environments (G-test: G (22) = 58.817, $P <$ 0.001). Likewise, *E. coli* plasmids significantly differed in antibiotic resistance classes, carbohydrate and nitrogen cycling traits, and plasmid mobility gene content across human and animal plasmid environments of ($P < 0.001$ for all tests).

We next removed the three most abundant genera (*Acinetobacte*r, *Escherichia*, and *Klebsiella*) from *Proteobacteria* and the most abundant genus *Staphylococcus* from *Firmicutes* and tested that the results held within these phyla. Amongst the 3,330 remaining *Proteobacteria* plasmids, overall COG composition, resistance genes, carbon and nitrogen utilization traits, and plasmid mobility genes still varied across the seven environments ($P < 0.001$ for all tests). Similarly, the same differences held across environments amongst the 1,028 remaining *Firmicute* plasmids.

DISCUSSION

Plasmids may confer traits that allow a bacterial host to adapt to its local environment. In support of this hypothesis, we identified a high diversity of accessory genes on plasmids from a large public database and found that the prevalence of these genes varied significantly by

environment. These cross-environment differences held for a variety of taxonomic subsets of the data, suggesting that these patterns are not entirely driven by database biases.

While previous studies have revealed differences in plasmid traits within an environment or among similar habitat types (Sentchilo et al., 2013; Luo et al., 2016; Brown Kav et al., 2020; Perez et al., 2020; Dewar et al., 2021), our analyses across seven broad environments provides a baseline of plasmid accessory trait distribution for future studies. Similar to traits encoded on the bacterial chromosome, we show that plasmid accessory trait distributions reflect their environment. For instance, the composition of CAZymes on plasmids varied across environments, just as patterns of CAZymes from whole genomes also vary. In particular, GH1 and GH3 – two of the most abundant GH families in bacterial genomes (Berlemont and Martiny, 2013) – were found relatively evenly on plasmids from all environments. In contrast, CAZymes that are highly abundant in human environments, e.g., GH5 and GH32 (Berlemont and Martiny, 2016) – are also more prevalent in plasmids from human environments.

Notably, plasmid accessory traits seem to form two larger groupings by environment, that is, accessory traits from human/animal/wastewater and those from plant/soil/marine/freshwater environments. This could be due to higher bacterial dispersal across some environments relative to others. For instance, microbial communities of wastewater environments reflect the microbiomes of human populations, through sewer systems and surface runoff (J. et al., 2015). Plasmids encoding clinically-relevant resistance genes and human pathogenic bacteria have previously been identified in wastewater environments, which have been suggested as hotspots for plasmid mediated horizontal gene transfer of resistance genes (Shannon et al., 2007; Rahube and Yost, 2010; Yang et al., 2013).

We also identified a high diversity of COG functions on plasmids that are typically associated with the bacterial chromosome as they are conserved genes found across diverse groups of bacteria (Galperin et al., 2021). These plasmid sequences may offer a "snapshots" of the HGT event between bacteria in time but how this connects with core versus flexible/accessory genome, and open versus closed state of pangenomes (all the genes present in a given species, across all isolates), is unclear (Brockhurst et al., 2019). However, by quantifying the diversity of plasmid accessory genes across different environments from publicly available data, we offer a glimpse of the possible selective pressures that may contribute to the maintenance of these genes on mobile genetic elements in bacterial communities.

It was previously thought that many environmental plasmids do not generally encode any obvious accessory genes, and for some of the environments, such as freshwater, this may be the case. For instance, 10 broad host range plasmids isolated from freshwater bacteria did not encode any accessory genes (BC J. et al., 2013). Here, the 101 freshwater plasmids also carried relatively few accessory genes (COG). Future studies of freshwater and other non-human associated environments would help to reveal whether this is a true biological pattern. Differences in plasmid traits across environments will also be driven by differences in host taxonomic composition if plasmids are at least partially conserved with their host phylogeny. Indeed, taxonomy at the phylum level appears to be a large factor in influencing trait distributions than the environment, at least for the five phyla with plasmids represented in all seven environments. Contrary to our hypothesis, this pattern held for both plasmid mobility and accessory traits. Thus, despite high levels of horizontal gene transfer of plasmids within phyla (Bonham et al., 2017; Redondo-Salvo et al., 2020), plasmid traits seem to be strongly influenced by their host's evolutionary history; only the composition of COG functions were significantly

affected by the environment after controlling for host phylum. This result is in alignment with recent work indicating that plasmids and their mobility traits are constrained by the evolutionary history of their hosts, even as some hosts are permissive to gene exchange beyond species/genus borders (Harrison and Brockhurst, 2012; Redondo-Salvo et al., 2020; Hall et al., 2021). We caution, however, that the analysis here gives an approximation; our ability to test the role of host taxonomy is limited by the uneven data available across phyla and environments. In the future, it would be useful to investigate the changing influence of taxonomy versus environment on plasmid traits as one moves from broader to finer taxonomic scales, where HGT may be more prevalent (Cohan, 2001; Doolittle and Papke, 2006).

Finally, in support of our last hypothesis, plasmid accessory traits varied in their association with the presence of MOB genes. Not surprisingly, resistance traits were most often found on potentially mobilizable plasmids (Partridge et al., 2018). The connection between plasmid traits and mobility also varied by environment; plasmid accessory traits appear more mobilizable in human environments relative to other environments (with the exception of N-genes in plant plasmids). This pattern may be due to culturing biases; however, stable temperatures and resources, along with high concentrations of bacteria cells, may make human environments especially favorable for HGT (Lerner et al., 2017). A caveat of our analysis is that the presence of MOB genes does not indicate the potential for plasmid transmissibility or actual HGT frequency (Sheppard et al., 2020). For instance, there may be unrecognized (non-MOB gene related) mechanisms of plasmid transfer, particularly in less well-studied environments. Indeed, nontransmissible plasmids appeared to be widely disseminated among members of a *Vibrionaceae* population, despite the lack of a clear mechanism for transmission (Xue et al., 2015).

CONCLUSION

A key aspect of plasmid diversity and evolution is the connection between plasmids, their bacterial hosts, and the environment. Our results suggest that plasmid-bound traits offer a substantial source of genetic diversity for bacterial adaptation to their environment, but more work is needed to directly link these mobile genetic elements to host adaptation in natural communities. While we focused on plasmid sequences from cultured microbes, this approach limited the data available, but links plasmid traits directly with its host, an advantage that culture-independent plasmidome studies do not have. Thus, implementing sequencing approaches like Hi-C (Lieberman-Aiden et al., 2009) into future plasmidome studies could also help address this disconnect between plasmids and their host, and has shown promise in some studies connecting the resistome and plasmidome to wastewater microbiomes (Stalder et al., 2019). Additionally, by combining current methods in novel ways such as cell enumeration and sorting with amplicon sequencing of microbial communities, plasmids fitness effects can be studied within microbial communities (Li et al., 2020), thereby helping to understand the adaptive role of plasmids. Further, efforts to culture and sequence plasmids from underrepresented clades and environments, such as *Cyanobacteria* and *Actinobacteria* and aquatic environments, would be enormously valuable as 75% of the plasmid sequences analyzed here were from *Proteobacteria*. Finally, an outstanding question is the role that plasmids have in microbial functioning at the community scale. Recent experiments demonstrate that manipulation of mobile genetic elements in communities may influence biogeochemical processes such as nitrogen cycling (Quistad et al., 2020), pointing to exciting directions for investigating how plasmid evolution influences the ecology of microbial communities.

ACKNOWLEDGMENTS

We thank Cynthia I. Rodriguez, Nicholas C. Scales, Alberto Barrón-Sandoval, and Kristin Barbour, and Joanna Zhao Wang for helpful insights and comments in drafts of the manuscript. This work was supported by a National Science Foundation Graduate Research Fellowship and UCI Chancellor's Club Fellowship to SSF, and by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research grants DE-SC0016410 and DE-SC0020382.

TABLES AND FIGURES

Table 2.1 Summary of plasmid genetic content across environments.

	Human	Animal	Wastewater	Plant	Soil	Marine	Freshwater	Totals
Plasmid sequences	5,122	2,316	227	1,089	657	213	101	9,725
Gene calls*	439,164	201,694	21,017	422,909	136,619	27,715	14,567	1,263,685
Mean plasmid size (kb)	76.3	79.3	88.4	426.9	215.5	130.1	146.6	-
No. of MOBs	2,686	1,077	154	633	325	96	54	5,025
% mobile**	50.5	43.7	59.5	52.5	45.7	41.3	47.5	-
No. of assigned COGs	227,768	104,664	11,742	319,138	89,239	17,062	9,548	779,161
No. of resistance genes	3,628	839	109	159	70	21	15	4,841
No. of CAZymes	3,740	1,716	158	9,107	2,217	339	263	17,540
No. of N-genes	354	210	46	2,601	556	141	42	3,950

*Excludes partial gene calls

** Percent of plasmids that contain at least one MOB gene

Table 2.2 PERMANOVA* testing the effects of environment and host phylum (taxonomy) by trait type.

Trait/factors	DF	SS	MS	Pseudo-F	*P*-value	ECV (%)
COG						
Environment	6	123.38	20.56	1.58	**0.044**	9.29
Host Phyla	3	203.98	67.99	5.23	**0.001**	40.12
Residuals	13	169.04	13.00			50.59
Resistance						
Environment	6	124.93	20.82	0.92	0.710	-
Host Phyla	3	113.40	37.80	1.67	**0.008**	13.45
Residuals	10	225.90	22.59			86.55
Carbon cycling						
Environment	6	763.41	127.23	1.05	0.295	-
Host Phyla	4	755.32	188.83	1.56	**0.001**	9.65
Residuals	17	2056.50	120.97			90.35
Nitrogen cycling						
Environment	6	174.59	29.10	0.94	0.733	-
Host Phyla	4	187.90	46.97	1.52	**0.002**	9.83
Residuals	15	464.23	30.95			90.17
Mobility						
Environment	6	35.14	5.86	0.85	0.780	-
Host Phyla	4	136.75	34.19	4.97	**0.001**	44.23
Residuals	16	110.18	6.89			55.77

*Significant *P*-values are boldface, indicates statistical significance with $P < 0.05$; *P*-values based on 999 permutations (lowest *P*-value possible 0.001); DF – degrees of freedom; SS – sum of squares; Pseudo-F – F value by permutation; ECV – estimated components of variation; Formula – Trait distance matrices ~ Environment + Host Phyla.

Table 2.3 Summary of traits associated with mobilizable* plasmids by environment.

Trait by plasmid MOB counts	Average (%)	Human	Animal	Wastewater	Plant	Soil	Marine	Freshwater
COGs								
Mobile		134195	48646	6652	133049	28232	6077	4821
Nontransmissible		93917	56133	5102	186109	61022	10993	4728
% Mobile	46.4	58.8	46.4	56.6	41.7	31.6	35.6	50.5
Resistance								
Mobile		2380	563	58	68	24	15	11
Nontransmissible		1248	276	51	91	46	6	4
% Mobile	64.4	65.6	67.1	53.2	42.8	34.3	71.4	73.3
CAZymes								
Mobile		3563	1227	181	3770	745	145	128
Nontransmissible		3127	1862	135	5643	1774	291	179
% Mobile	45.5	53.3	39.7	57.3	40.1	29.6	33.3	41.7
N-gene families								
Mobile		114	66	11	1394	137	37	14
Nontransmissible		240	144	35	1207	419	104	28
% Mobile	44.9	32.2	31.4	23.9	53.6	24.6	26.2	33.3

*Mobilizable = plasmids having at least one MOB gene identified.

Table 2.4 Summary of COG categories associated with mobilizable plasmids (the presence of MOB genes).

COG category	Mobilizable plasmid	Nontransmissible plasmid	% COG on mobilizable plasmids
A	6	18	25.00
B	2	6	25.00
C	12,634	17,886	41.40
D	12,758	10,936	53.84
E	21,187	31,014	40.59
F	3,052	5,024	37.79
G	24,115	31,102	43.67
H	11,021	15,003	42.35
I	11,726	19,011	38.15
J	7,553	8,972	45.71
K	30,081	36,726	45.03
L	34,257	23,976	58.83
M	15,628	21,040	42.62
N	4,311	8,295	34.20
O	12,555	11,226	52.79
P	16,794	21,650	43.68
Q	4,483	7,820	36.44
T	15,596	24,192	39.20
U	18,453	10,947	62.77
V	20,377	18,773	52.05
W	367	623	37.07
X	54,582	55,283	49.68
Z	22	11	66.67
R+S	29,990	38,077	44.06

Table 2.5 Summary of resistance genes associated with mobilizable plasmids (presence of MOB genes).

Resistance class	Mobilizable	Nontransmissible	% Class associated with mobilizable plasmids
Aminoglycosides	61	53	53.51
Arsenic	43	57	43.00
Beta-lactams	1587	852	65.07
Biocide and metal	12	29	29.27
Cadmium	1	2	33.33
Cationic antimicrobial peptides	8	6	57.14
Copper	19	12	61.29
Drug and biocide	96	67	58.90
Fluoroquinolones	27	9	75.00
Fosfomycin	8	3	72.73
Glycopeptides	6	9	40.00
Iron	0	1	0
Lead	96	43	69.06
Lipopeptides	2	1	66.67
Mercury	659	302	68.57
Metal	256	136	65.31
Metronidazole	1	7	12.50
MLS	37	26	58.73
Multi-biocide	1	6	14.29
Multi-drug	5	5	50.00
Naphthoquinone	136	41	76.84
Nickel	9	11	45.00
Phenicol	6	9	40.00
Rifampin	2	11	15.38
Sulfonamides	25	8	75.76
Tellurium	1	0	100.00
Tetracyclines	8	9	47.06
Trimethoprim	7	7	50.00

Table 2.6 Summary of CAZymes by family associated with mobilizable plasmids (presence of MOB genes).

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
AA10	2	35	5.41
AA3	82	92	47.13
AA3_2	78	64	54.93
AA4	4	7	36.36
AA5_2	0	5	0.00
AA6	1	8	11.11
AA7	15	1	93.75
CBM0	1	0	100.00
CBM12	1	9	10.00
CBM13	24	55	30.38
CBM16	2	1	66.67
CBM2	5	88	5.38
CBM20	2	12	14.29
CBM21	0	1	0.00
CBM22	1	3	25.00
CBM23	0	2	0.00
CBM26	1	1	50.00
CBM27	0	1	0.00
CBM3	0	4	0.00
CBM32	14	108	11.48
CBM34	3	3	50.00
CBM35	5	1	83.33
CBM4	1	3	25.00
CBM40	0	1	0.00
CBM41	0	3	0.00
CBM42	0	2	0.00
CBM48	157	411	27.64
CBM5	49	227	17.75
CBM50	10	41	19.61
CBM51	2	6	25.00
CBM54	0	1	0.00
CBM56	0	1	0.00
CBM57	3	0	100.00
CBM6	1	2	33.33
CBM61	0	3	0.00
CBM63	0	1	0.00
CBM66	1	1	50.00
CBM67	3	1	75.00
CBM85	0	1	0.00
CBM9	0	2	0.00

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
CE0	10	6	62.50
CE1	56	29	65.88
CE11	3	3	50.00
CE12	0	4	0.00
CE14	29	15	65.91
CE15	7	1	87.50
CE16	0	3	0.00
CE2	0	1	0.00
CE3	0	3	0.00
CE4	133	201	39.82
CE5	17	18	48.57
CE6	0	2	0.00
CE7	5	8	38.46
CE8	7	86	7.53
CE9	10	23	30.30
GH0	183	344	34.72
GH1	72	112	39.13
GH10	2	18	10.00
GH102	0	7	0.00
GH103	49	102	32.45
GH104	4	6	40.00
GH105	43	42	50.59
GH106	8	4	66.67
GH108	11	19	36.67
GH109	158	42	79.00
GH110	0	1	0.00
GH112	1	0	100.00
GH113	1	1	50.00
GH114	1	2	33.33
GH115	1	2	33.33
GH117	2	0	100.00
GH12	6	15	28.57
GH123	0	1	0.00
GH125	1	2	33.33
GH126	1	0	100.00
GH127	29	33	46.77
GH128	0	1	0.00
GH13	5	6	45.45
GH13_10	0	1	0.00
GH13_11	0	2	0.00
GH13_14	0	1	0.00
GH13_16	49	124	28.32

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
GH13_18	20	8	71.43
GH13_19	0	2	0.00
GH13_20	2	1	66.67
GH13_21	1	0	100.00
GH13_23	7	21	25.00
GH13_26	38	145	20.77
GH13_27	0	1	0.00
GH13_28	1	0	100.00
GH13_29	1	5	16.67
GH13_3	1	7	12.50
GH13_30	0	2	0.00
GH13_31	30	16	65.22
GH13_32	0	1	0.00
GH13_33	1	0	100.00
GH13_36	0	1	0.00
GH13_4	0	2	0.00
GH13_5	1	3	25.00
GH13_8	1	0	100.00
GH130	11	5	68.75
GH133	1	3	25.00
GH135	2	80	2.44
GH136	4	1	80.00
GH137	1	1	50.00
GH139	1	0	100.00
GH140	6	0	100.00
GH141	0	1	0.00
GH142	1	0	100.00
GH143	1	0	100.00
GH144	11	3	78.57
GH145	2	0	100.00
GH146	0	1	0.00
GH148	24	0	100.00
GH15	47	55	46.08
GH151	0	5	0.00
GH154	51	22	69.86
GH159	0	1	0.00
GH16	80	42	65.57
GH17	0	3	0.00
GH18	2	25	7.41
GH19	27	15	64.29
GH2	176	102	63.31
GH20	50	51	49.50

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
GH23	2640	676	79.61
GH24	20	135	12.90
GH25	45	40	52.94
GH26	26	12	68.42
GH27	1	2	33.33
GH28	54	223	19.49
GH29	45	60	42.86
GH3	124	120	50.82
GH30	0	1	0.00
GH30_1	0	1	0.00
GH30_2	0	1	0.00
GH30_5	0	2	0.00
GH31	11	24	31.43
GH32	147	188	43.88
GH33	49	54	47.57
GH35	6	8	42.86
GH36	64	33	65.98
GH37	2	87	2.25
GH38	35	38	47.95
GH39	9	35	20.45
GH4	119	134	47.04
GH42	22	41	34.92
GH43	0	1	0.00
GH43_1	2	1	66.67
GH43_10	1	2	33.33
GH43_11	1	3	25.00
GH43_12	6	2	75.00
GH43_17	1	1	50.00
GH43_18	2	0	100.00
GH43_22	0	3	0.00
GH43_24	1	0	100.00
GH43_26	4	2	66.67
GH43_27	0	1	0.00
GH43_29	3	0	100.00
GH43_30	0	1	0.00
GH43_32	0	1	0.00
GH43_5	1	0	100.00
GH43_8	1	0	100.00
GH43_9	2	0	100.00
GH44	0	1	0.00
GH46	1	1	50.00

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
GH5	3	4	42.86
GH5_1	2	21	8.70
GH5_12	0	1	0.00
GH5_13	3	0	100.00
GH5_2	0	1	0.00
GH5_28	1	0	100.00
GH5_39	0	1	0.00
GH5_4	0	1	0.00
GH5_44	0	1	0.00
GH5_46	1	3	25.00
GH5_48	7	7	50.00
GH5_5	2	80	2.44
GH5_7	2	0	100.00
GH5_8	2	0	100.00
GH50	1	1	50.00
GH51	38	49	43.68
GH52	0	1	0.00
GH53	0	21	0.00
GH55	0	1	0.00
GH57	1	0	100.00
GH6	0	3	0.00
GH63	40	22	64.52
GH65	13	7	65.00
GH66	0	1	0.00
GH67	2	1	66.67
GH68	14	5	73.68
GH70	2	2	50.00
GH73	163	174	48.37
GH74	0	2	0.00
GH75	0	1	0.00
GH76	1	2	33.33
GH77	32	120	21.05
GH78	22	48	31.43
GH79	3	0	100.00
GH8	12	23	34.29
GH81	0	1	0.00
GH84	7	8	46.67
GH85	8	0	100.00
GH86	1	0	100.00
GH88	38	52	42.22
GH89	2	0	100.00

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
GH9	3	1	75.00
GH91	9	25	26.47
GH92	7	3	70.00
GH93	0	1	0.00
GH94	27	56	32.53
GH95	3	13	18.75
GH97	2	1	66.67
GT0	125	309	28.80
GT1	167	211	44.18
GT10	0	6	0.00
GT102	1	5	16.67
GT103	0	4	0.00
GT107	52	4	92.86
GT11	1	18	5.26
GT14	0	3	0.00
GT17	0	9	0.00
GT19	1	2	33.33
GT2	606	1014	37.41
GT20	31	187	14.22
GT21	8	4	66.67
GT23	17	2	89.47
GT25	6	15	28.57
GT26	19	58	24.68
GT28	2	12	14.29
GT3	1	1	50.00
GT30	1	3	25.00
GT32	2	88	2.22
GT35	25	36	40.98
GT38	1	0	100.00
GT39	0	1	0.00
GT4	626	883	41.48
GT41	19	109	14.84
GT42	1	0	100.00
GT44	30	121	19.87
GT45	1	0	100.00
GT5	39	96	28.89
GT51	94	82	53.41
GT53	3	0	100.00
GT56	0	1	0.00
GT6	1	0	100.00
GT60	0	3	0.00

CAZyme ID	Mobilizable	Nontransmissible	% CAZymes associated with mobilizable plasmids
GT7	0	6	0.00
GT70	0	1	0.00
GT73	15	2	88.24
GT75	0	1	0.00
GT8	11	13	45.83
GT81	1	2	33.33
GT83	24	116	17.14
GT87	2	1	66.67
GT89	0	2	0.00
GT9	5	126	3.82
GT99	0	2	0.00
PL0	3	92	3.16
PL1	9	6	60.00
PL1_2	25	0	100.00
PL1_4	0	1	0.00
PL1_6	2	1	66.67
PL10_1	2	1	66.67
PL11	0	1	0.00
PL12	1	3	25.00
PL15_1	0	3	0.00
PL22_1	0	3	0.00
PL22_2	1	0	100.00
PL26	0	1	0.00
PL3	0	3	0.00
PL3_1	26	1	96.30
PL3_2	9	2	81.82
PL33_1	31	21	59.62
PL4_1	0	2	0.00
PL5	0	4	0.00
PL7	0	1	0.00
PL8	2	4	33.33
PL9	0	5	0.00
PL9_1	30	3	90.91
PL9_2	1	20	4.76

Table 2.7 Summary of N-gene families associated with mobilizable plasmids (presence of MOB genes).

N-gene family ID	Mobilizable	Nontransmissible	% N-genes associated with mobilizable plasmids
amoC_B	0	1	0.00
ansB	7	17	29.17
asnB	1	1	50.00
gdh_K00260	4	3	57.14
gdh_K00261	11	9	55.00
gdh_K00262	120	143	45.63
gdh_K15371	5	6	45.45
glnA	597	770	43.67
glsA	16	13	55.17
gs_K00264	10	1	90.91
gs_K00265	16	28	36.36
gs_K00266	115	192	37.46
napA	9	18	33.33
napC	12	15	44.44
narB	23	10	69.70
narC	1	7	12.50
narG	1	6	14.29
narI	5	18	21.74
narJ	2	22	8.33
narY	0	1	0.00
narZ	5	6	45.45
nasA	7	28	20.00
nifD	55	4	93.22
nifH	8	1	88.89
nifK	95	17	84.82
nifW	3	0	100.00
nirA	18	53	25.35
nirB	8	7	53.33
nirD	109	97	52.91
nirK	55	106	34.16
nirS	110	162	40.44
nmo	87	135	39.19
norB	0	1	0.00
norC	6	14	30.00
nosZ	118	167	41.40
NR	121	67	64.36
nrfC	8	15	34.78
ureA	0	1	0.00
ureB	1	2	33.33
ureC	4	13	23.53

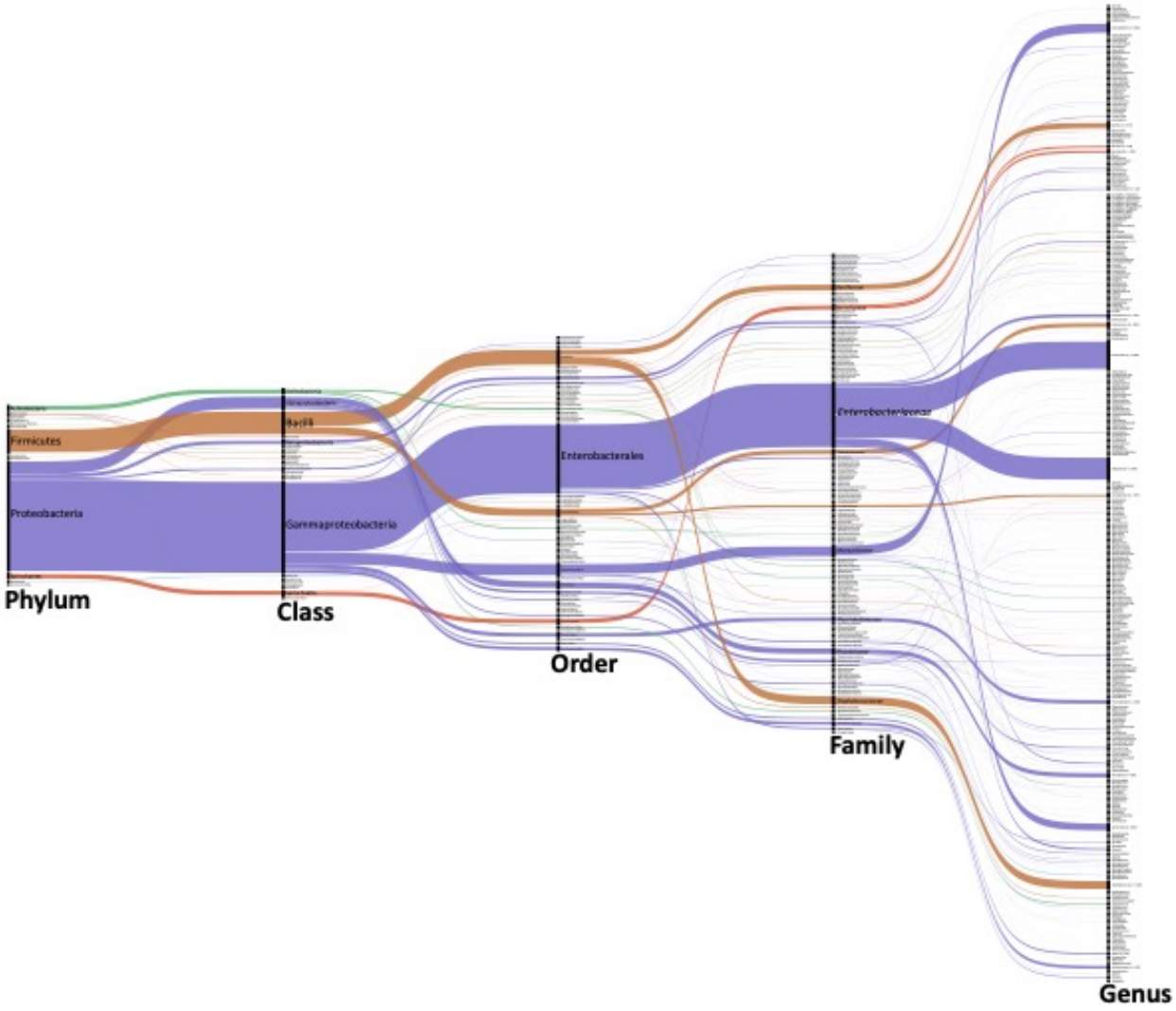

Figure 2.1 Representation of plasmid host bacteria in the PLSDB dataset across different taxonomic classifications. The more frequent the plasmid host bacterial taxa is represented in the dataset, the thicker the line weight. Note – the line color is by Phylum level taxonomic classification for simplicity.

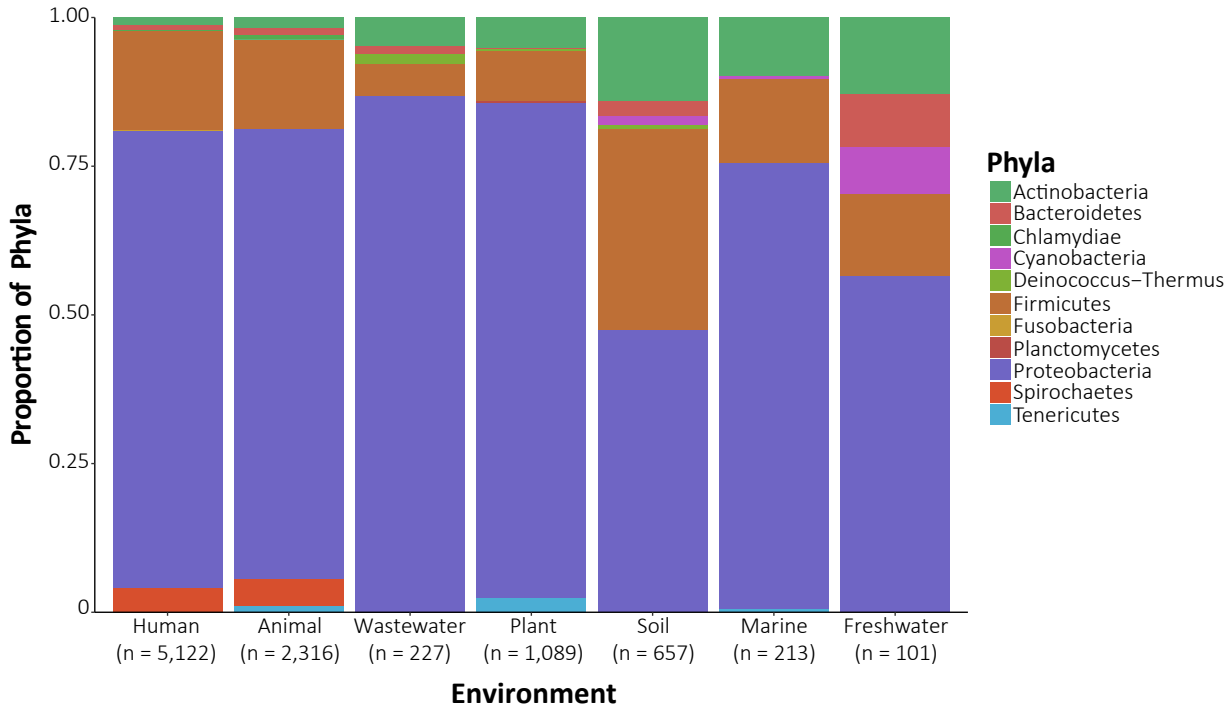

Figure 2.2 Taxonomic representation of plasmid host by environment in the PLSDB dataset. Note: plasmid host-phyla with < 1 % relative abundance are not shown, e.g., plasmids from *Elusimicrobia* (n=3).

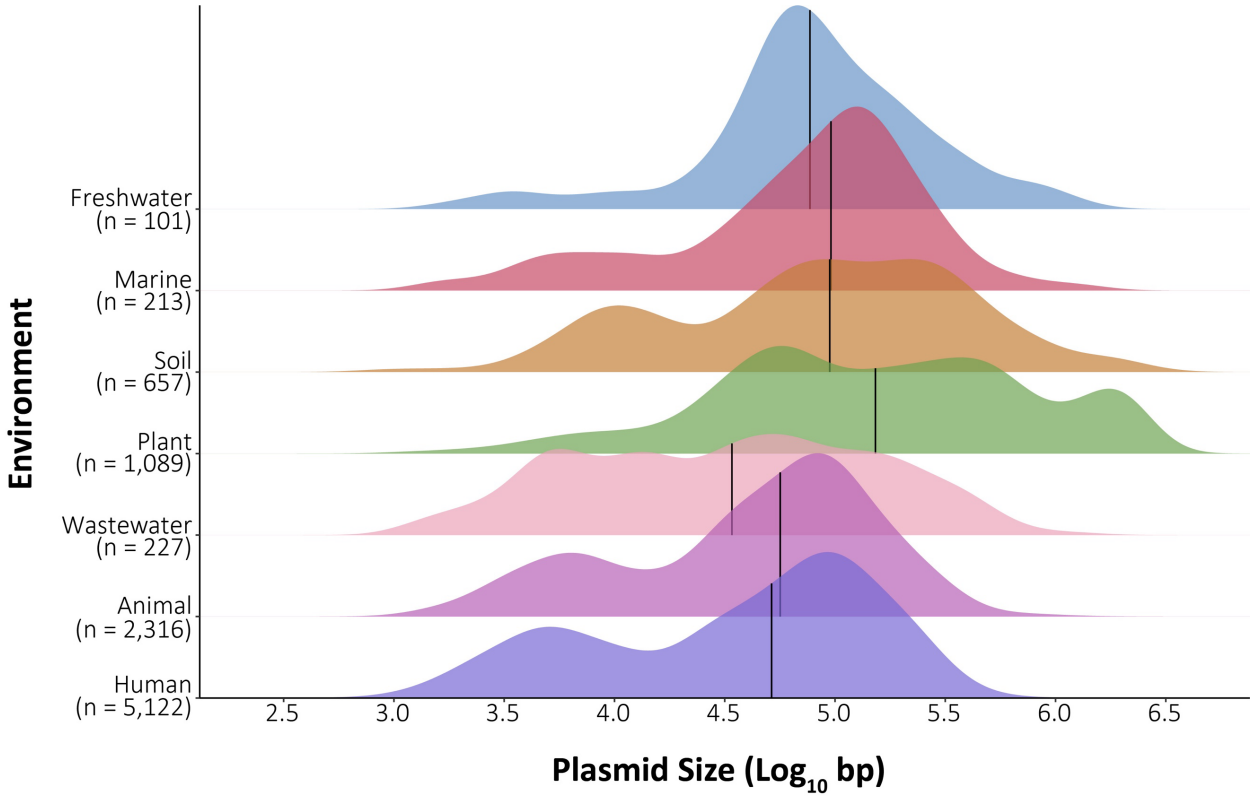

Figure 2.3 Plasmid size distributions vary significantly by environment. The density plots of plasmid nucleotide lengths by environment. The mean plasmid sizes are indicated by vertical black lines. Untransformed mean plasmid sizes are: human (76.3 kb), animal (79.3 kb), wastewater (88.4 kb), plant (426.9 kb), soil (215.5 kb), marine (130.1 kb), and freshwater (146.6 kb).

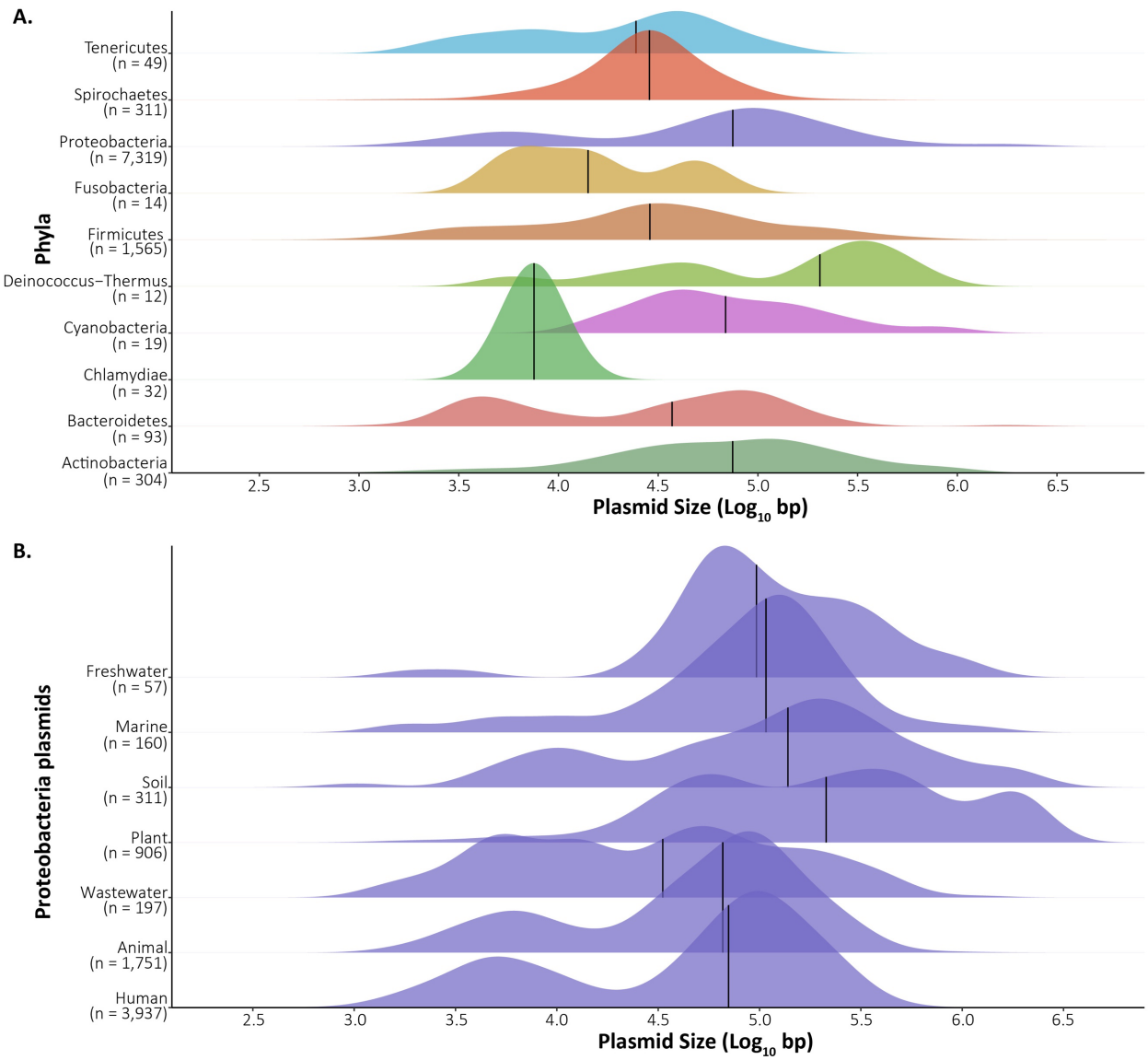

Figure 2.4 Plasmid size distributions vary significantly by host taxonomy and environment. The density plots of $\log_{10}$ transformed plasmid nucleotide lengths across 10 phyla (A) and across environments within Proteobacteria (B). The mean plasmid sizes are represented by vertical black lines at the center of respective density distributions. Plasmid size distributions for phyla having < 10 representatives are not shown. Untransformed mean plasmid sizes are as follows: *Actinobacteria* (140.1 kb), *Bacteroidetes* (73.1 kb), *Chlamydiae* (7.5 kb), *Cyanobacteria* (128 kb), *Deinococcus-Thermus* (225 kb), *Firmicutes* (68 kb), *Fusobacteria* (21.2 kb), *Proteobacteria* (146 kb), *Spirochaetes* (33.9 kb), *Tenericutes* (32.7 kb). Likewise, mean *Proteobacteria* plasmids by environment sizes are as follows: human (85.3 kb), animal (85.4 kb), wastewater (81.5 kb), plant (492.4 kb), soil (284.1 kb), marine (139.6 kb), and freshwater (190.4 kb).

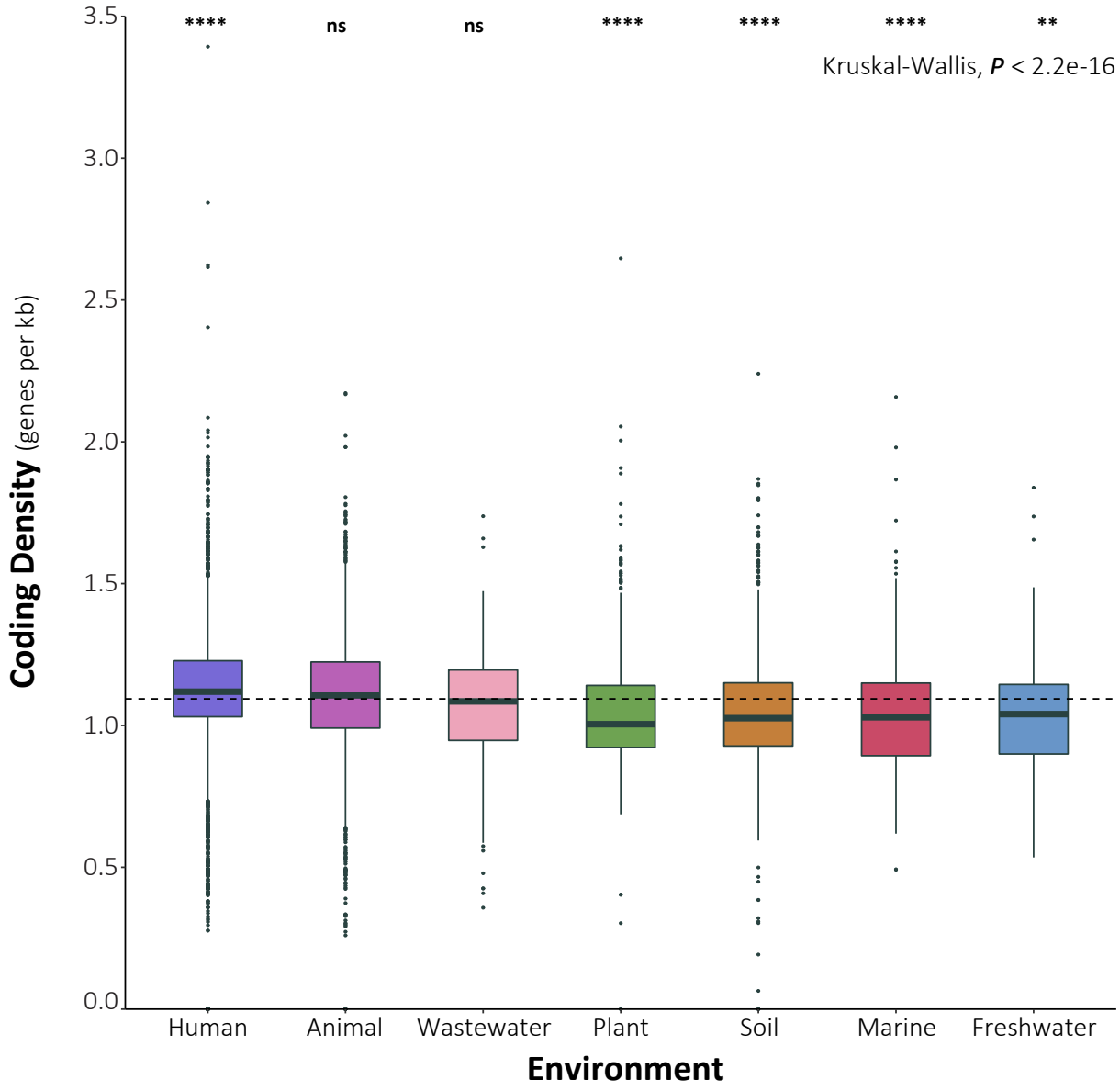

Figure 2.5 Plasmid coding density by environment. The dashed line represents the grand mean coding density (1.07) across environments. Symbols above each boxplot represent significant *P*-values for coding density for each environment compared to the grand mean as determined by Kruskal-Wallis tests.

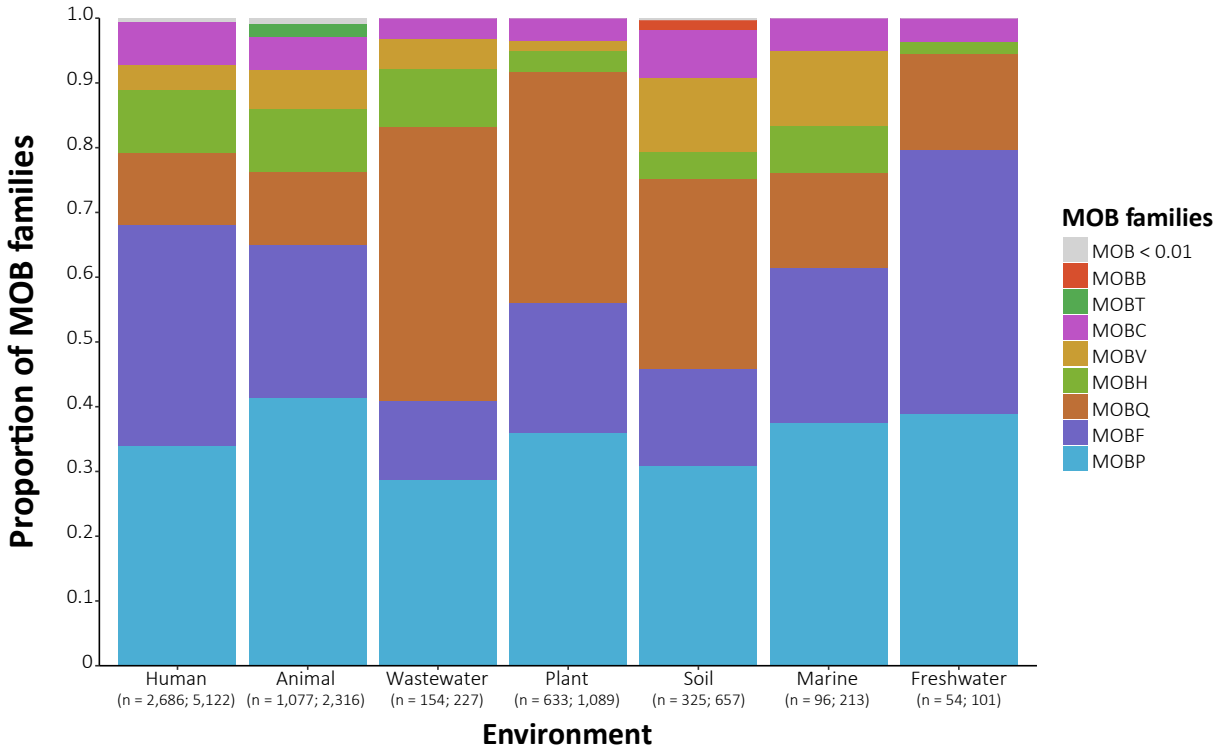

Figure 2.6 Proportion of MOB family of relaxases by environment. The values under each plasmid environment reflect the number of MOB relaxases identified, followed by the number of plasmids within an environment. MOB families represented by < 1% relative abundance within an environment are shaded in gray.

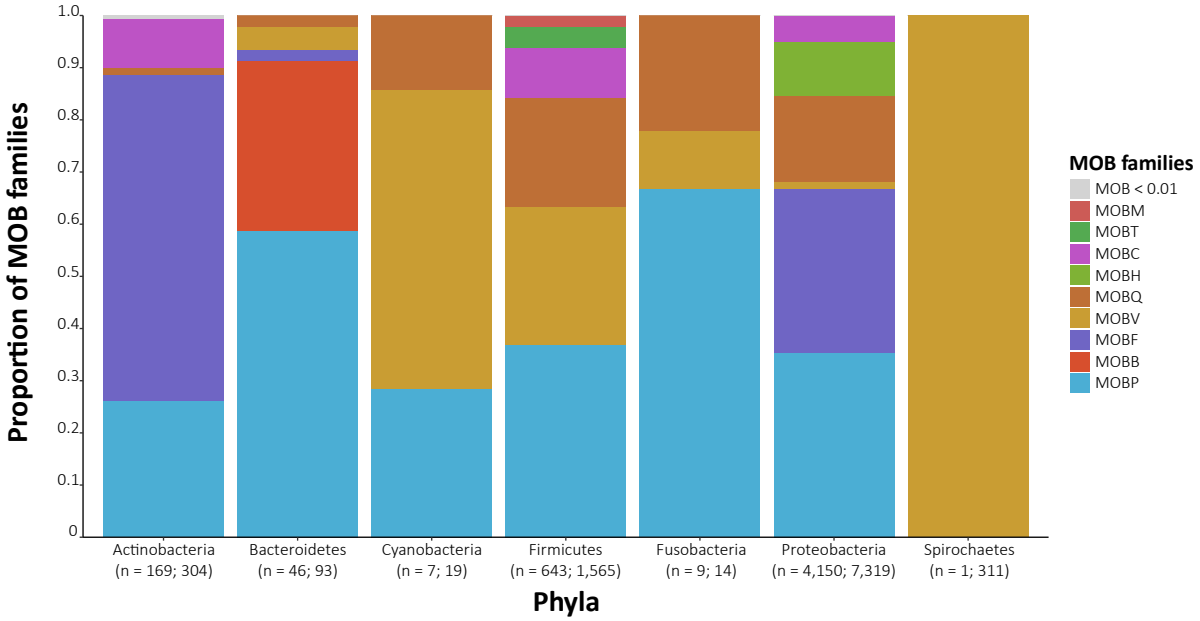

Figure 2.7 The proportion of MOB family relaxases by plasmid host-taxonomy. The values under each plasmid host phylum (n) reflect the number of MOB relaxases identified, followed by the total number of plasmids within a phylum. MOB families representing < 1 % relative abundance within an environment are shaded in gray.

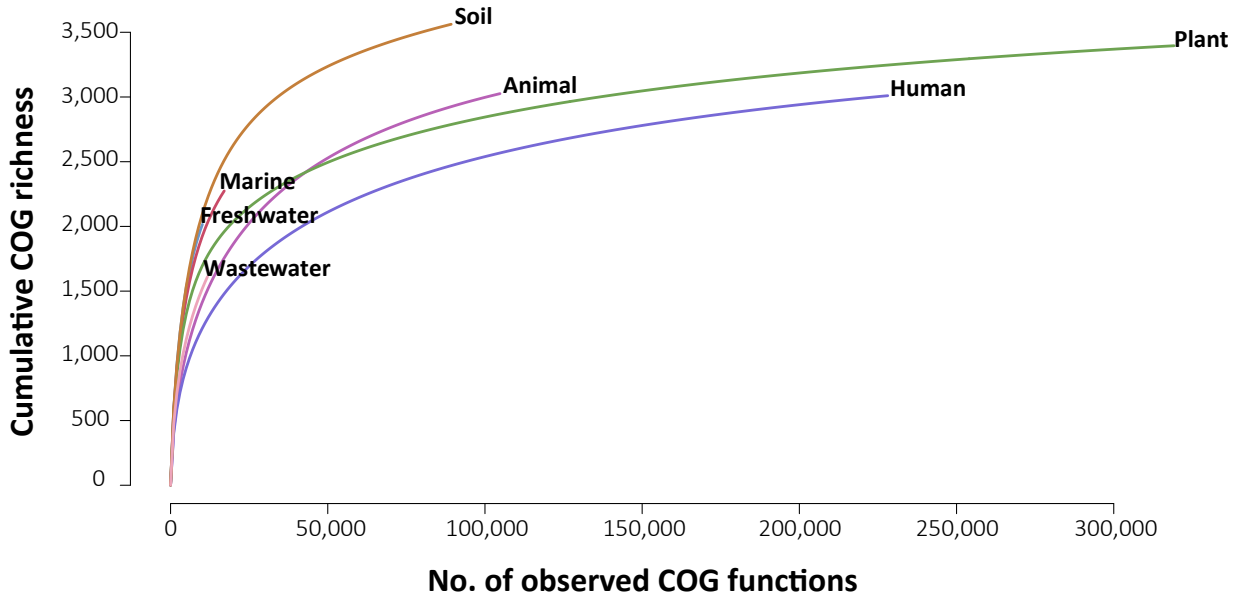

Figure 2.8 Cumulative COG richness of plasmids by environment. Unique COG function accessions were iteratively and randomly subsampled (n = 1,000 times) across each environment.

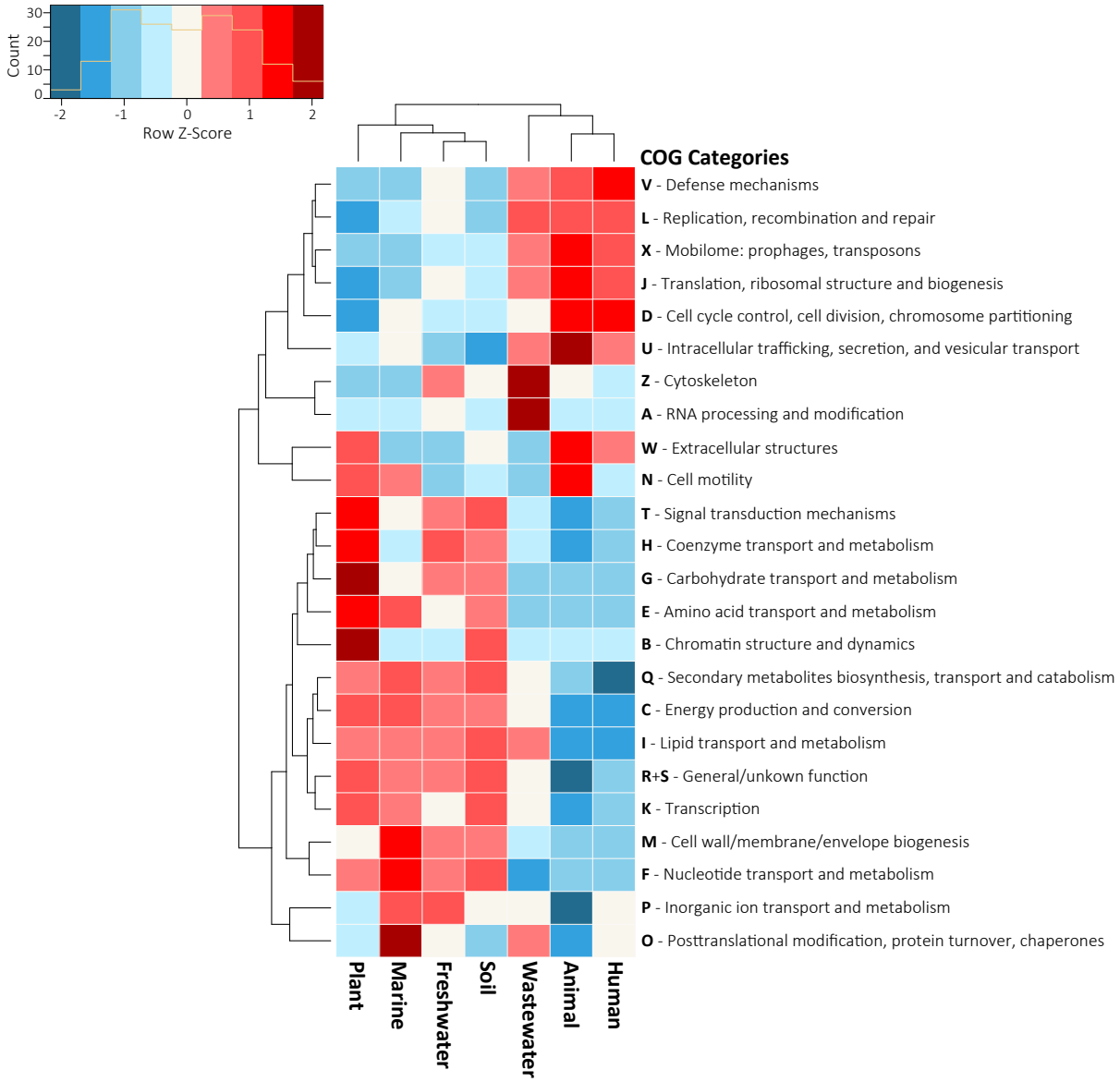

Figure 2.9 Plasmid COG functions by environment. The normalized frequencies of COG functions by environment at broader COG category designations. To standardize for uneven plasmid sequences across environments, COG function counts were first converted into proportional abundances within an environment after removal of COG functions < 6 identified across all environments. COG abundance across environments were then normalized using Z-scores. COG categories with above mean (white tiles) values are represented by tiles shaded in red, while those with lower than mean values are shaded in blues.

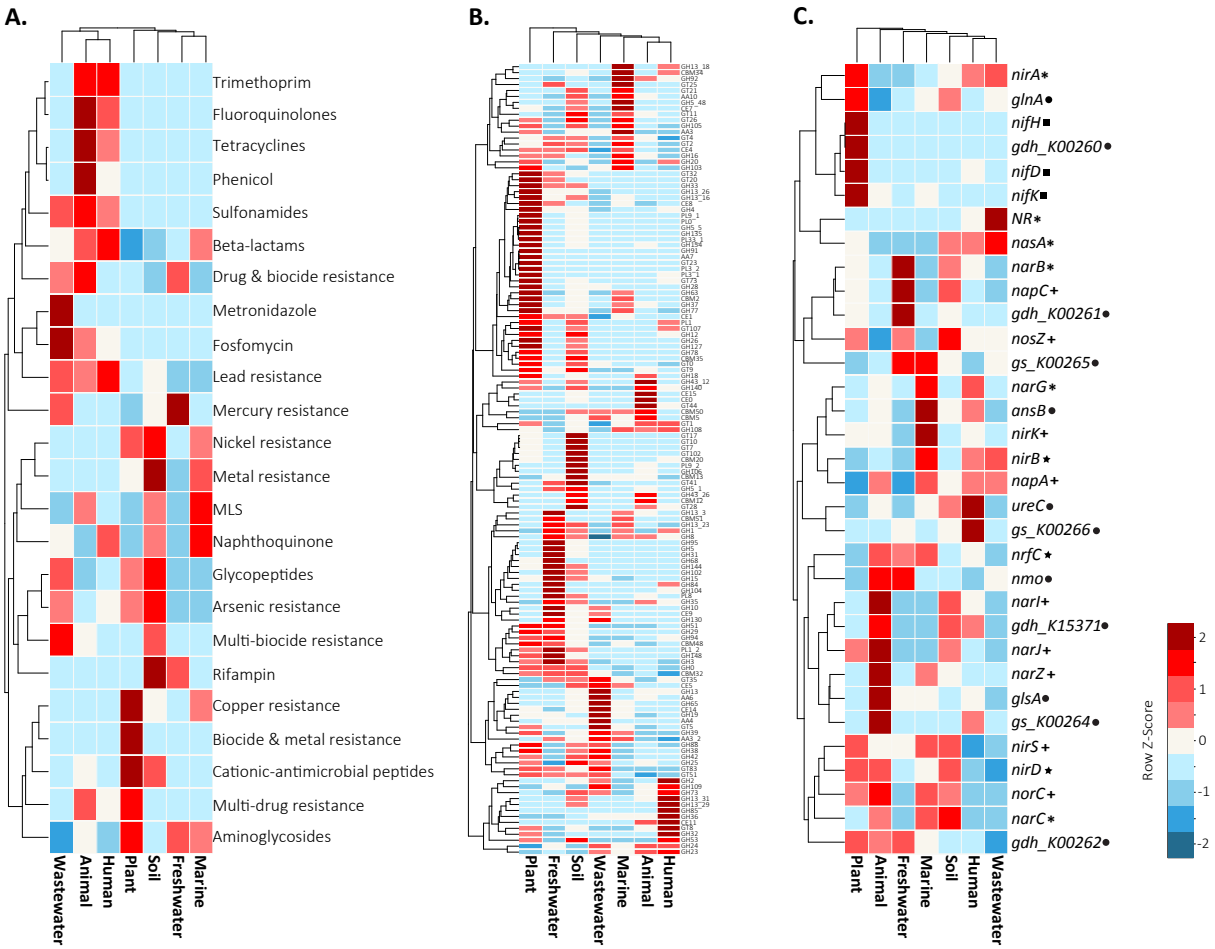

Figure 2.10 Plasmid resistance, carbon-degradation and nitrogen cycling traits by environment. Normalized frequencies of resistance genes by classes (**A**), CAZyme family types (**B**) and N-gene families (**C**) by environment. The resistance, CAZyme and N-gene families with above mean (white tiles) values are represented by tiles shaded in red, while those with lower than mean values are shaded in blues. Symbols next to N-gene family names represent corresponding N-cycling pathways: asterisks = assimilatory nitrate reduction pathways, plus = denitrification, stars = dissimilatory nitrate reduction, squares = nitrogen fixation, circles = organic degradation and synthesis.

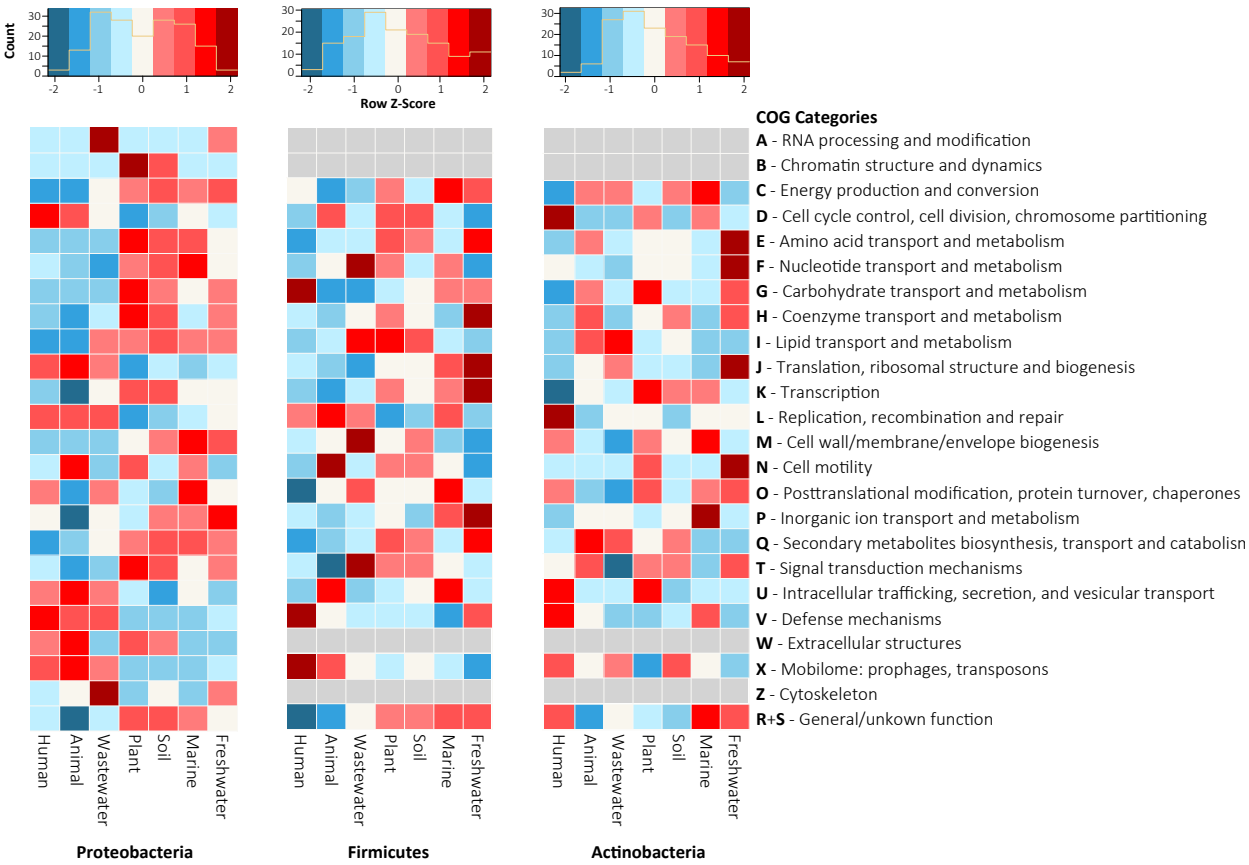

Figure 2.11 Plasmid COG categories by host phylum and environment. The standardized and normalized COG functions by environment at broader COG category designations are shown for the top three most abundant plasmid host phyla.

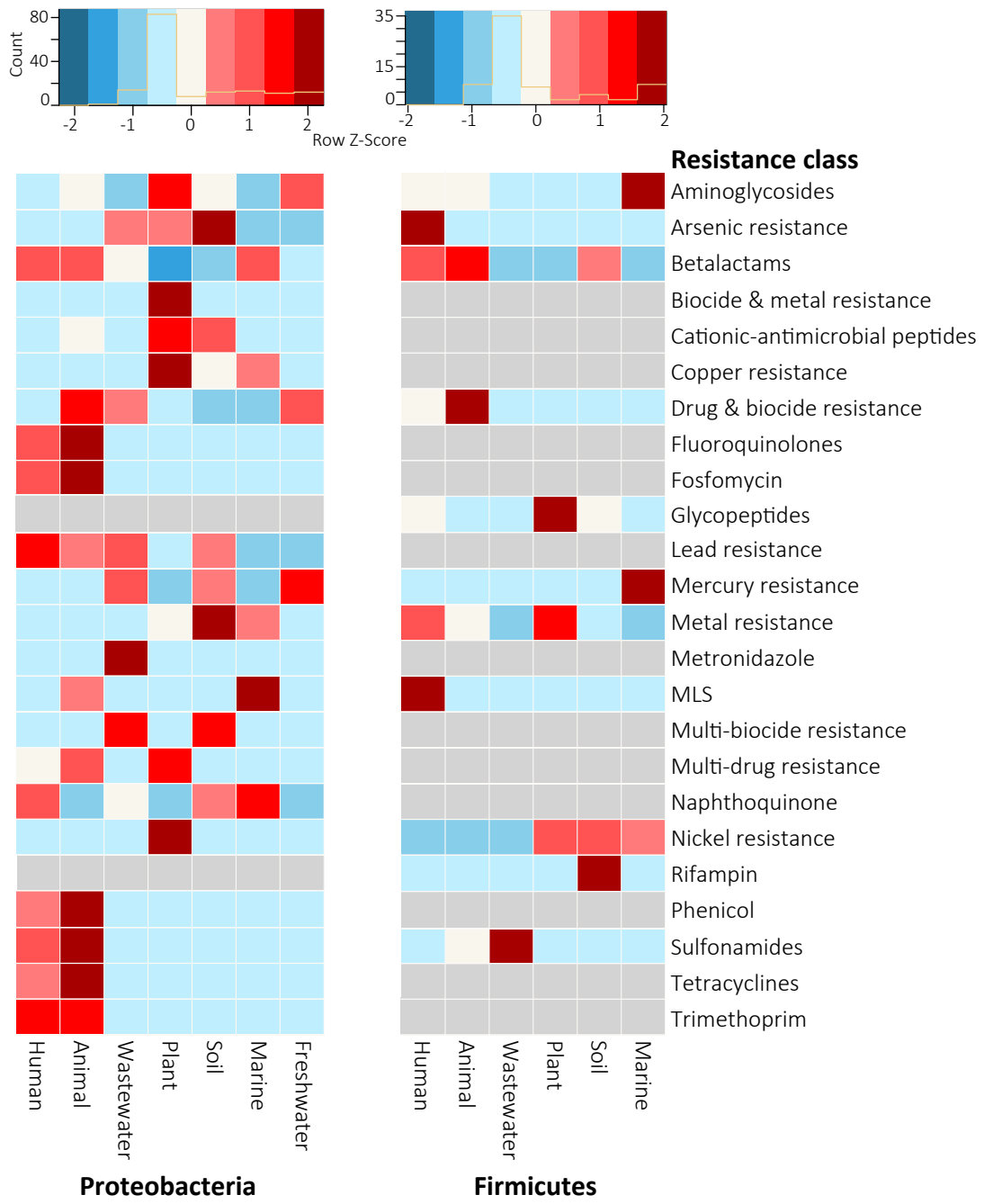

Figure 2.12 Plasmid resistance traits by host phylum and environment. The resistance genes are grouped by resistance class for plasmids of Proteobacteria and Firmicutes host. Resistance genes with above mean (white tiles) values are represented by tiles shaded in red, while those with lower than mean values are shaded in blues. Note: resistance genes of Firmicutes from freshwater environments are not shown due to low frequencies (< 15 genes). Gray tiles represent undetected resistance classes where no corresponding genes were identified.

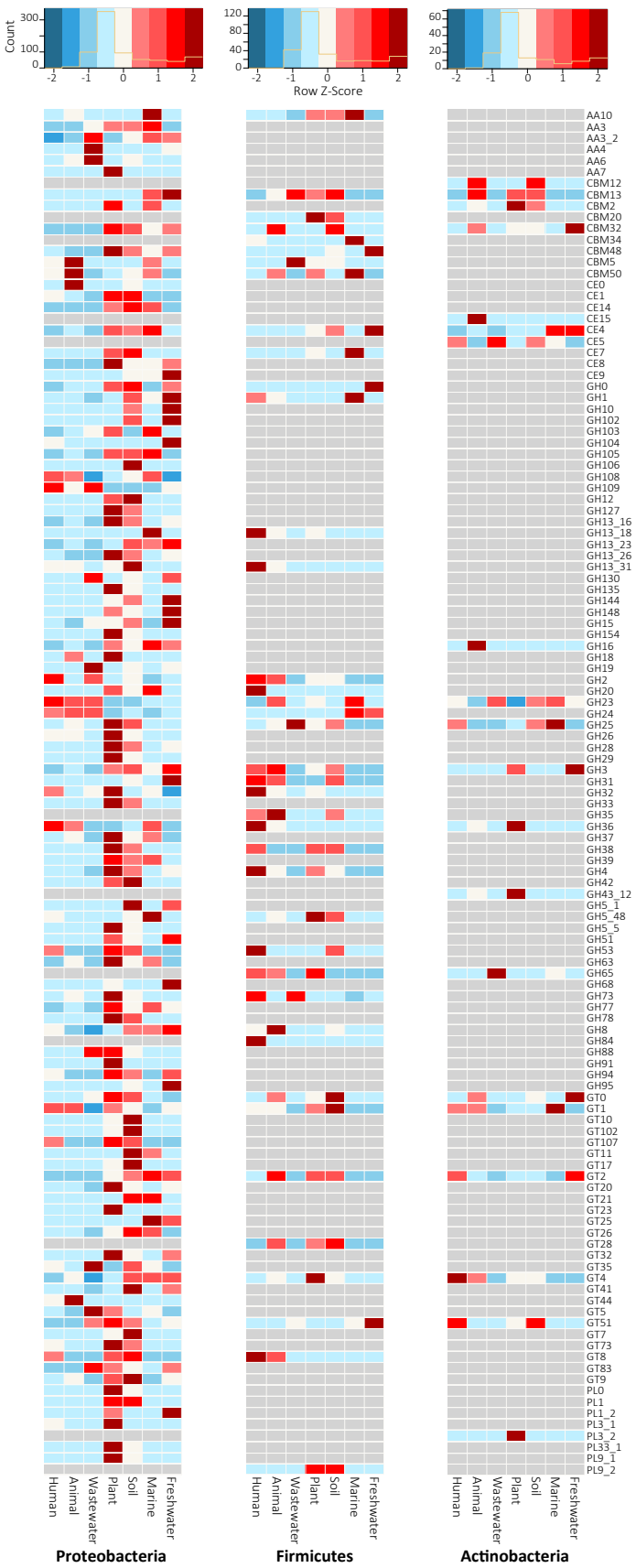

Figure 2.13 Plasmid carbon degrading traits by host phylum and environment. CAZymes are ordered alphabetically for plasmids of Proteobacteria, Firmicutes, and Actinobacteria. CAZymes with above mean (white tiles) values are represented by tiles shaded in red, while those with lower than mean values are shaded in blues. Gray tiles represent undetected instances where no corresponding CAZymes were identified within that host phylum.

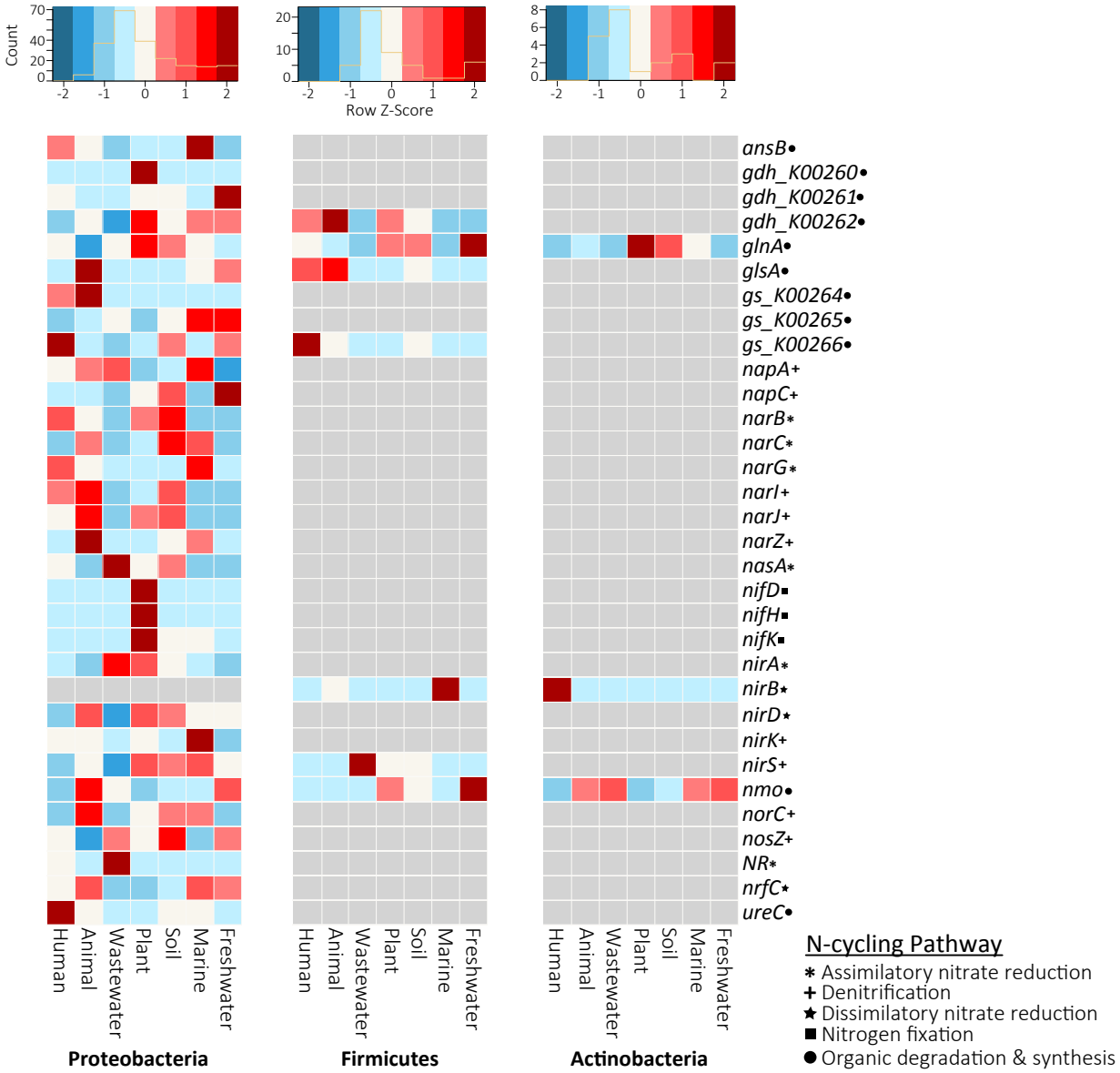

Figure 2.14 Plasmid nitrogen cycling traits by host phylum and environment. The N-gene families are ordered alphabetically for plasmids of Proteobacteria, Firmicutes, and Actinobacteria. Symbols at the right of the N-gene family names correspond to N-Cycling pathways. N-gene families with above mean (white tiles) values are represented by tiles shaded in red, while those with lower than mean values are shaded in blues. Gray tiles represent undetected instances where no corresponding N-gene families were identified within that host phylum.

CHAPTER 3

Plasmids contribute to trait variation within the soil taxon *Curtobacterium*

ABSTRACT

Our knowledge of the diversity and distribution of most bacteria has focused on core genomic signatures. Much less is known about how accessory genes encoded on mobile genetic elements (MGE), contribute to bacterial genetic diversity (with the exception of antibiotic resistance and virulence genes). Here, we investigated the role of MGE and their associated traits in *Curtobacterium*, an abundant bacterium found on decaying plant litter in southern California. Using both short and long-read sequencing data, we performed hybrid assemblies to produce 22 completely assembled genomes of *Curtobacterium* isolated across a climate gradient. We quantified the trait variation attributed to MGEs, and particularly to plasmids, identified in *Curtobacterium* genomes from our study and those deposited in public databases, and then determined whether the plasmid traits were habitat- versus ecotype- specific. More than half of the *Curtobacterium* genomes contained at least one complete plasmid, which were predominantly found within the genomes of bacterial hosts from one clade/subclade. About 1.5% of the total clusters of orthologous genes identified in the genomes (n = 27,974) were found on plasmids and included genes involved in carbon degradation (glycoside hydrolases) and inorganic/organic nitrogen processing (*narB*). Most plasmids also encoded at least one mobility gene, suggesting that traits associated with these plasmids are potentially mobilizable within the bacterial communities where the strains were isolated. Lastly, the breadth of trait diversity associated with plasmids suggests the potential for MGE to facilitate *Curtobacterium*'s adaptation to environmental change.

INTRODUCTION

Mobile genetic elements (MGE) including plasmids, prophages, and integrative conjugative elements, potentially allow a cell to rapidly adapt to sudden environmental changes (Frost et al., 2005; Heuer and Smalla, 2012; Siguier et al., 2014). MGE are broadly involved in three mechanisms of horizontal gene transfer (HGT) - transformation, conjugation, or transduction (Frost et al., 2005) – and thereby mediate movements of accessory traits within genomes or between bacterial cells. Through their contributions to HGT, MGE are key players in bacterial diversification (Croucher et al., 2014; Acman et al., 2020). Indeed, the extent to which MGE influence the traits of their bacterial hosts is exemplified by the global dissemination of genes that confer resistance against antibiotics (Carattoli, 2009; Carattoli, 2013; J. et al., 2021).

Plasmids are particularly successful in sharing genetic information via bacterial conjugation in microbial communities (Norman et al., 2009) and can allow bacteria to share genes with other bacteria over large taxonomic distances (Klümper et al., 2015). However, plasmids can also be vertically transmitted to daughter cells during host cell replication (Giraldo, 2003), which complicates quantifying the frequency of plasmid-mediated HGT within microbial communities. Plasmid diversity and composition in a community may thus be influenced by a combination of host phylogeny and selection by the local biotic and abiotic environment (Heuer et al. 2008 and 2012).

Understanding the extent to which plasmids contribute to genomic diversity within natural microbial communities is important to investigating how MGE may facilitate adaptation to environmental changes (Touchon et al., 2020; Conrad et al., 2022), Chapter 2). Thus, a first step in understanding the role of plasmids in environmental communities is to quantify the genetic diversity of plasmids and the traits they encode (Perez et al., 2020), and how they relate to the host bacterial taxa (Stalder et al., 2019). However, the evolutionary pressures that

determine the location (chromosomal or plasmid) of bacterial traits are not fully understood, but evidence suggests that the distribution of genes between plasmid and chromosomes is not random. In particular, antibiotic resistance genes appear commonly on plasmids (Eberhard, 1989; Fatoba et al., 2022; Meng et al., 2022), and other plasmid genes allow bacteria to survive extreme environments (Ryo et al., 2006; Romaniuk et al., 2018; Hanka et al., 2022). By contrast, genes essential to cell survival are thought to be located on the chromosome (Lehtinen et al., 2021). Still, many genes – presumably accessory genes – are found in both chromosomes and plasmids.

Here, we explore the contribution of plasmids to the genomic diversity of *Curtobacterium*, a cosmopolitan genus of bacteria associated with plants and soil (Chase et al., 2016). This genus shifts in abundance and composition in response to environmental changes (Matulich et al., 2015; Chase et al., 2017; Martiny et al., 2017; Chase et al., 2018). Furthermore, *Curtobacterium* are capable of responding rapidly to environmental shifts via *de novo* mutations (Chase et al., 2021). However, whether *Curtobacterium* also evolves rapidly via trait acquisition mediated by MGE such as plasmids remains unknown. Because *Curtobacterium* are relatively easy to culture (Chase et al., 2016), investigating the distribution of plasmids within this host genus is tractable through long-read sequencing. Previous studies of *Curtobacterium* species associated with plant disease have found that their plasmids carry virulence traits (Chen et al., 2020), but otherwise, the abundance and diversity of plasmids in this genus remains unknown.

Using long read sequencing of cultured isolates, we identified and sequenced new plasmids within this genus and compared them to publicly available plasmid sequences. To investigate the drivers influencing plasmid diversity, we first asked whether *Curtobacterium* plasmids and their traits are phylogenetically conserved within a host ecotype and/or vary by

environment? We hypothesized that extensive vertical transmission within this genus would result in the phylogenetic conservation of plasmids within a host ecotype - that is, previously-defined genetic clusters that correspond to ecologically-relevant phenotypes (Chase et al., 2018). Thus, we expected that some regions of *Curtobacterium* plasmids would be conserved, as detected in plasmids associated with *Enterobacterales* (Redondo-Salvo et al., 2020). Alternatively, high rates of horizontal transfer of the plasmids might break up any phylogenetic signal. In this case, local selection could result in an association between plasmid traits and the environment from which a particular strain was isolated without phylogenetic conservation.

Second, we asked: What traits do *Curtobacterium* plasmids encode, and do they differ from those found on the chromosome? Because *Curtobacterium* is associated with natural soil environments, we did not expect them to carry genes associated with xenobiotics or metal resistance, except perhaps when isolated from agricultural environments. More generally, we hypothesized that functional genes that are more prevalent on plasmids than the chromosome might indicate the potential for these traits to be beneficial across species of *Curtobacterium* and facilitate rapid adaptation to the perturbations in the soil environment.

METHODS

Strains and reference sequences. We analyzed three collections of *Curtobacterium* strains. First, 60 *Curtobacterium* from senescent plant litter samples taken along an elevation gradient in Southern California and previously sequenced on an Illumina platform (Chase et al., 2018; Glassman et al., 2018; Glassman and Martiny, 2018).

Second, we sequenced four novel *Curtobacterium* strains obtained from Napa Valley, CA. and the Loma Ridge Global Change Experiment field site (Finks et al., 2021), The strains, 54.4B, S.11, S.43, and Napa-18, were obtained from, 50–100 mg of ground plant litter samples,

each spread over media prepared with grass litter leachate and agar. The media was prepared using 100 mg of ground grass litter collected from the Loma Ridge field site in July 2017 and suspended in 1.0L of deionized water under constant stirring for approximately 12 hours. The leachate was then filtered through cheese cloth to remove the bulk of plant litter particulate and adjusted to pH 7.0 ± 0.1 and 1.5 g of bacteriological agar (BD Difco, Maryland, USA) added per 1.0 L of leachate. The media was autoclaved at 121°C for 15 minutes, and the filter-sterilized antifungal agent, Cycloheximide (Acros Organics, New Jersey, USA) added to molten media at a final concentration of 0.1 mg/ml. Plant litter samples plated on litter media were monitored daily for growth, and colonies were screened for phenotypic characteristics attributed to *Curtobacterium* strains (Evtushenko and Takeuchi, 2006). Selected colonies were streaked onto LB agar plates until pure. Strains were then inoculated into 5.0 mL of LB and incubated at room temperature and 225 rpm until turbid growth appeared. All strains were stored in a 25 % glycerol solution at -80 °C.

Finally, complete *Curtobacterium* plasmid and chromosome sequences matching the search criteria '*Curtobacterium'* and 'Chromosome', or 'Plasmid', were retrieved from the NCBI GenBank and RefSeq databases on March 31, 2022 and analyzed for the presence of plasmids[9]. In total, the *Curtobacterium* strains included in analysis were isolated from seven distinct environments, include 25 % from agricultural crops, while the majority of strains (58 %) originate from environments along a climate gradient isolated from plant litter, the top layer of soil[10].

[9] supplemental_materials_1.xlsx
[10] supplemental_materials_1.xlsx

DNA preparation and sequencing of strains. DNA extraction of novel *Curtobacterium* strains with the addition of 200 µg/ml of lysozyme from chicken egg white (Sigma-Aldrich, Missouri, USA) were performed using the Wizard Genomic DNA purification kit from Promega (Wisconsin, USA). Short read sequencing of these DNA extractions were run on and Illumina NovaSeq 6000 instrument with an S4 flow cell aiming for 800M reads, and 300 cycles paired end (Illumina Inc., California, USA). Libraries were prepared using the Nextera Flex kit (now the Illumina Prep library kit) using the low volume protocol. The samples were pooled based on qubit reads to equal quantities (assuming similar fragment length).

To add to the complete plasmid and chromosome sequences retrieved from NCBI databases, twenty-two *Curtobacterium* strains representing five distinct ecotypes across a climate gradient in southern California (Chase et al., 2018) were selected for sequencing on an Oxford Nanopore Technologies (ONT) platform. DNA extractions of these strains was performed using Qiagen Blood and Cell Culture DNA Mini Kit (Qiagen, Hilden, Germany). This extraction method generates high molecular weight gDNA (> 60 kb), free of small DNA contamination, that is suited to ONT sequencing. The extracted DNA was quality assessed via Nanodrop (Thermo Fisher; Massachusetts, USA) and quantified by Qubit (BioTek; Vermont, USA). ONT sequencing libraries were multiplexed and run on three different MiniION devices with R9.4.1 flow cells (300 Mbp; The SeqCenter Team formerly The Microbial Genome Sequencing Center, Pennsylvania, USA) and high accuracy basecaller Guppy v5.0.16.

Sequence assemblies. *De novo* 'hybrid' assemblies of ONT and Illumina sequenced *Curtobacterium* strains were performed with quality checked short and long reads using the default settings of Unicycler version 0.4.8 (Ryan R Wick et al., 2017). We chose a 'hybrid' assembly (combining long and short read sequencing data) approach for the *Curtobacterium* that

were ONT sequenced, as many long reads can exceed the length of repeats in bacterial genome, which are also a characteristic of many types of MGE, and short reads can improve accuracy, thus this method results in genomic assemblies with highly accurate structure and sequences (Ryan R. Wick et al., 2017). For all other sequenced *Curtobacterium*, *de novo* short-read only assemblies were obtained using Unicycler with default settings. Both ONT and Illumina sequencing data was checked for quality using FastQC version 0.11.9[11] and visualized using MultiQC version 1.9 (Ewels et al., 2016). Adapter sequences and low-quality reads (PRED < 8) and sequences less than < 2 kbp were removed from ONT data per previously described methods for long read preparation (Ryan R. Wick et al., 2017). Briefly, adaptors were trimmed and chimeric sequences removed from ONT sequencing data using Porechop version 0.2.4[12]. Some chimeras were detected in our read sets, which can occur with multiplex Nanopore sequencing (Xu et al., 2018) and were removed during read trimming with Porechop[13]. For Illumina generated data, adapter sequences and low quality (PHRED < 30) reads were removed using FastP version 0.20.0 (Chen et al., 2018). The read quality for ONT and Illumina prepared reads, was then reassessed using FastQC and MultiQC. Unicycler generated assembly graphs were visualized in Bandage version 0.8.1 (Wick et al., 2015). To assess the completeness of genome assemblies, the web interface of Quast[14] (Gurevich et al., 2013), was used determine assembly contiguity and completeness (e.g., N50, BUSCO, and %GC).

Plasmid and trait characterization. All assemblies were first run through the RAST server – a rapid annotation tool using subsystems technology (Aziz et al., 2008) to identify

[11] https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[12] https://github.com/rrwick/Porechop#trim-adapters-from-read-ends
[13] https://github.com/rrwick/Porechop
[14] http://quast.sourceforge.net/

general features on chromosomal and plasmid (e.g., Type IV Secretion Systems, genes encoding plasmid replicases and partitioning during host replication) assemblies. The RASTtk annotation scheme included: Prodigal (Hyatt et al., 2010) to make gene calls, the option to automatically fix errors (e.g., gene candidates overlapping with RNAs), and replication disabled. Complete putative plasmid and chromosome sequences from hybrid assemblies were searched for MOB family relaxases, which are essential for conjugative DNA processing (Garcillán-Barcia et al., 2009) using MobScan[15] (Garcillán-Barcia et al., 2020). All putative plasmid and chromosome assembly sequences were screened for phage genes using PHASTER (Arndt et al., 2016), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated genes (Cas) systems using CRISPRCasFinder (Couvin et al., 2018), Integrating and Conjugating Elements/Conjugative Elements (also called conjugative transposons) using blastp search of protein sequences downloaded on August, 31, 2021 from ICEberg version 2.0 database (Liu et al., 2019), using threshold E-value lower than 10^{-10}, a minimum alignment coverage of 50 % and with at least 70 % identity were selected, and the name of the best hit extracted.

Complete putative plasmid and chromosome assemblies were imported into Anvio version 7.0 (Eren et al., 2021) for comparative genomics analysis. Plasmid and chromosome sequences were analyzed separately within Anvio, and gene calls were made with Prodigal and searched via COG20, Clusters of Orthologous Groups of genes/proteins (Galperin et al., 2021) and Pfam version 33.1 (Mistry et al., 2021) databases with DIAMOND v0.9.14 (Buchfink et al., 2015) in sensitive mode. Clustering analysis of plasmid and chromosome amino acid sequence similarity search results were performed in Anvio using MCL algorithm (van Dongen and Abreu-Goodger, 2012), excluding partial gene calls, default settings for minbit heuristic and

[15] https://castillo.dicom.unican.es/mobscan_about/

MCL inflation parameter, and a minimum gene cluster occurrence of 1. To organize gene clusters for chromosome/plasmids, Euclidean distance and ward linkage parameters were selected. Functional enrichment analysis to determine which plasmid COG (functions were enriched by environment (for this with > 5 putative plasmids represented) and ecotype (clades where > 5 plasmids are represented) was performed in Anvio (Shaiber et al., 2020). Briefly, Anvio associates every gene cluster with a function, functions are assigned at the gene level. Ideally all the genes in a single gene cluster would be annotated with same function but if there are multiple functions associated with a gene cluster, the most frequent function, or the one with the largest number of genes in the cluster is selected. Only functions with q-values (adjusted p-values for multiple comparisons - Rao test) below 0.05 significance level were reported and thus considered as enriched functions. All COG functions, Pfam hits, and corresponding gene calls were exported as tables from Anvio and merged before importing into R (R Core Team (2020), 2020) for further analysis.

Additionally, we analyzed plasmids for genes involved in carbohydrate and nitrogen utilization. To do this, the standalone version of dbCAN2, release 07-31-2019 (Zhang et al., 2018), and the NCycDB – with curated nitrogen (N) gene family sequences (encompassing seven N-cycling pathways) release 2019 (Tu et al., 2019) was used. For carbohydrate utilization trait analyses, genes encoding for Carbohydrate Active Enzymes (CAZymes) hits in the dbCAN2 database were included if two or more of the three search tools (HMMER, DIAMOND, Hotpep) matched in their identification of the same CAZyme family. For instances where a single gene call returned multiple matches to CAZyme families, only the annotations from the first domain hit were included in downstream analyses. For N-gene family hits, BLASTp searches of plasmid gene calls having E-values 10^{-5} and >50 % query coverages were included in the analyses.

Phylogenomic and statistical analysis. To identify conserved regions and plasmid/chromosome similarity, the Average Nucleotide Identity (ANI) was calculated for complete plasmid and chromosome assemblies in Anvio via PyANI (Pritchard et al., 2016) and using BLASTN+ to align 1020 nt fragments of each assembly (Goris et al., 2007; Michael and Ramon, 2009), and visualized using the anvio-display-pan feature of Anvio interactive interface. To infer evolutionary associations between chromosomal assemblies, single-copy core genes (SCGs) were identified within Anvio and nucleotide positions that were gap characters in more than 50% of the sequences were removed from concatenated SCGs using trimAl[16]. For phylogenomic analysis, IQ-TREE (Nguyen et al., 2015; Minh et al., 2020) with the 'WAG' (Whelan and Goldman, 2001) general matrix model was used to infer a maximum likelihood tree. Newick tree files were exported and visualized using iTOL version 5 (Letunic and Bork, 2021) to output cladograms or phylogenomic trees. To infer influence of evolutionary history (ecotype/clade designation from phylogenomic analysis) for putative plasmids, ANI full percentage identity (ANI normalized over the assembly length) was used, since no single-copy conserved genes were identified for all plasmid sequences. Similarity scores were then exported from Anvio and analyzed further in R and PRIMERe.

To test whether *Curtobacterium* plasmids and chromosomes ANI similarities varied by ecotype and/or environment from which they were isolated, permutational multivariate analysis of variance PERMANOVA (permutations $n = 999$ with unrestricted permutations of raw data using type III sums of squares) were performed in PRIMER-e version 6 (Clarke and Gorley, 2006b; Anderson et al., 2008b) with ecotype and environment designated as fixed factors. For these tests, rare ecotypes (i.e., *Curtobacterium* chromosomes outside clade I or V) and

[16] https://github.com/inab/trimal

environments (e.g., *Curtobacterium* isolated from algae or unknown origins) were lumped together and categorized as 'other'. The estimated percentage of variance explained was determined by dividing terms with significant p-values by the sum of the estimates of components of variation given as output from PRIMER-e and multiplying by 100. To test whether plasmids and their host chromosome ANI were correlated, a RELATE test (Anderson et al., 2008b) using Spearman correlation was performed in PRIMER-e.

To determine if COG function composition varied by replicon type (plasmid or chromosome), operation functional units (OFU = COG counts by replicon type) were standardized by the assembly length, such that the proportional abundances of COG for a given assembly summed to one. Trait abundances by replicon type were then normalized using Z-scores in R v3.6.3 (R Core Team (2020), 2020). This procedure served to weight each trait similarly, rather than proportional to its abundance. To compare COG composition across replicons, we then calculated the Euclidean distance of the standardized and normalized OFU tables using the *vegdist* function of the 'vegan' package in R (Oksanen et al., 2019), and PERMANOVA and RELATE tests performed as previously mentioned for ANI analysis, but with OFU distance matrices. Ordinations of all similarity (ANI) and dissimilarity (COG) matrices were performed using non-metric multidimensional scaling approach with the *metaMDS* function of the 'vegan' and visualized using 'ggplot2' packages in R (Wickham, 2009). All other figures were constructed using 'ggplot2'[17] in R. To determine whether the proportions of *Curtobacterium* plasmid COGs were different among replicon types and environments, Log-Likelihood-Ratio (G-test) test of independence using the *GTest* function

[17] https://ggplot2.tidyverse.org

(with no correction) of the 'DescTools'[18] v0.99.44 package in R were performed. G-tests were performed on contingency tables of non-standardize trait counts with rare traits (traits counts < 6 across all environments) removed. To compare size distributions of replicons, Kruskal-Wallis tests on the nucleotide length (bp) of chromosome and plasmid sequences grouped by environment was performed using the 'stats' package of R. All other plots were generated using 'ggplot2' packages in R.

RESULTS

Using long-read sequencing to produce newly assembled genomes, we uncovered a total of 26 plasmids from 66.7 % of the 39 *Curtobacterium* strains investigated. Phylogenomic analysis of 916 single-copy conserved genes on the chromosomes revealed that nearly all the host strains belonged to five distinct ecotypes (based on clade designations); only two strains, AA3 and BH2-1-1, did not group within previously described clades (Figure 3.1A). The mean plasmid size and GC content of *Curtobacterium* plasmids was approximately 136 kb and 64.4 %, respectively. In contrast, the mean chromosome size and GC content was approximately 3.6 Mb and 71.3 %. The distribution of plasmid sizes did not vary significantly across *Curtobacterium* ecotypes (Kruskal-Wallis: H (5) = 8.963, P = 0.111, Figure 3.1B, supplemental_materials_1.xlsx). In contrast, the distribution of plasmid sizes varied significantly across *Curtobacterium* environments (Kruskal-Wallis: H (7) = 15.529, P = 0.029, Figure 3.1B, supplemental_materials_1.xlsx). Notably, several *Curtobacterium* strains carried multiple plasmids. For instance, Desert-03 had two plasmids, one of which (pD03b) is a small (1,579 bp), high-copy (> 8,000 X short read coverage) plasmid.

[18] https://andrisignorell.github.io/DescTools/

Curtobacterium plasmids showed little evidence for phylogenetic conservation by host ecotype, as the amount of overlap in genetic similarity was minimal among the plasmids. We examined putative plasmid sequences for conserved genomic regions (> 1020 bp) by host ecotype and found that only 5 *Curtobacterium* plasmids shared more than 65 % ANI, and all of these fell within host ecotype I (Figure 3.2). Not surprisingly, then, plasmid ANI was not correlated with host chromosome ANI (RELATE: r = 0.119, $P = 0.11$). Notably, the five plasmids from ecotype I with > 65 % ANI similarity all were isolated from hosts collected from agricultural environments.

Next, we investigated the traits that *Curtobacterium* plasmids encode, and whether they differed from those found on the chromosome. The 26 *Curtobacterium* plasmids encoded more than 4,000 genes that clustered into 2,396 gene clusters. Some of these clusters contained genes for plasmid features necessary for conjugative, cell-to-cell DNA transfer (e.g., *trwC*), and for plasmid partitioning to daughter cells during host replication and division (e.g., *parA/B/G*). The gene *lsr2*, a putative histone-like protein (important for protecting DNA and regulating gene expression), was shared across 17 of the 26 plasmids. We also identified multiple genes encoding putative CAZymes, ~ 20 % (n = 179) of which function were identified as glycoside hydrolases. In contrast, genes encoding functions important for nitrogen cycling, such as *narB*, were less common, being found only on plasmids pD35 and pS16 from desert and salton-sea strains (Figure 3.3). Most of the singleton gene clusters (genes shared by only one plasmid) were associated with mega-plasmids (> 500 kb), and in particular, with pD40 that was assembled from a desert *Curtobacterium* (ecotype V) strain.

In contrast to the low ANI similarity, many plasmids shared similar functional (COG) categories. This trait variation could be explained in part by both phylogenetic similarity

(ecotype) and the environment from which the host was isolated. The composition of COG functions on the plasmids was influenced by both the ecotype and environment ($P < 0.01$; Table 3.1; Figure 3.4A), explaining 19 % and 37 % of the variation, respectively. Notably, a similar amount of variation was also explained by ecotype and environment for COG composition on the chromosomal replicon (17 % and 38 %, respectively; Figure 3.4B). In particular, carbohydrate transport and metabolism (G) were relatively more abundant in plasmids isolated from a grassland, algae, and an alpine forest (Figure 3.5). Although, overall COG functions for carbohydrate transport and metabolism, were more abundant on chromosome of *Curtobacterium* compared to plasmids (Figure 3.6). In contrast, COG functions for post-translational modification, protein turnover, and chaperones (O) were highly abundant in plasmids from a desert scrubland relative to those from other environments. Interestingly, COG functions for coenzyme transport and metabolism (H), cell motility (N), and replication, recombination and repair (L), and translation, ribosomal structure and biogenesis (J) were prevalent across *Curtobacterium* plasmids from agriculture, desert, and salton-sea environments.

Finally, *Curtobacterium* plasmids encoded a different distribution of traits than their host chromosome. Overall, the plasmids genes broadly grouped into 22 broader COG categories, despite making up only 3 % of the total coding genetic content (n = 137,723 gene calls) across the genomes included in this dataset (Figure 3.6). Some COG functions were more prevalent to plasmids (G (21) = 1203.2, $P < 0.001$; Figure 3.6). For instance, the COG categories for replication, recombination, and repair (L) and cell motility (N) were relatively more abundant on plasmids (9.6 % and 7.9 %) than on chromosomes (4.2 % and 1.7 %). In contrast, COG categories for amino acid transport and metabolism (E) and lipid transport and metabolism (I)

were relatively more abundant on chromosomes than plasmids (8.6 % and 4.6 % versus 1.9 % and 2.7 %, respectively).

DISCUSSION

Plasmids contribute ecologically relevant diversity to the abundant and ubiquitous surface soil bacterium, *Curtobacterium*. The majority of strains investigated in this study carried from one to three plasmids, making up 1.4% of all the traits encoded on the genome. However, this percentage varied widely depending on the strain; the largest plasmid (pD40) contributed nearly 15% of the genome content of a strain.

Overall, there was limited phylogenetic similarity between the plasmids, even for quite closely-related strains isolated from the same location. Thus, contrary to our hypothesis, there was limited the phylogenetic conservation of plasmids within a host ecotype. This result suggests that horizontal transmission of plasmids within ecotypes is common in this genus and breaks up the signal of vertical transmission, at least at the genetic resolution sampled here. However, at the functional level, ecotype explained approximately twice as much of the total variation in COG composition on plasmids as the isolation environment, similar to the amount explained on the chromosome. This result indicates that historical (phylogenetic) context has a similar impact on broad trait content for both plasmid and chromosomal replicons. On the face of it, these results seem to conflict. However, *Curtobacterium* plasmids pools may be somewhat restricted to an ecotype, but within an ecotype, the clade's specific plasmids may undergo a high amount of horizontal transmission. Similarly, an analysis of over 10,000 plasmids within the order of *Enterobacterales* revealed that although most plasmids have host ranges beyond the species barrier, they also form discrete clusters or plasmid taxonomic units (Redondo-Salvo et al., 2020).

Curtobacterium plasmids also encoded a diversity of traits that are not a random subset of chromosomal traits. As expected, we did not detect many antibiotic or metal resistance genes on the plasmids. Instead, *Curtobacterium* plasmids were enriched (relative to the chromosome) in genes involved in functions for replication, recombination, and repair and cell motility. Together with the variation in plasmids found within an ecotype, this result suggests that these traits may come under relatively rapid selection and provide a benefit for maintaining a horizontally transferred plasmid.

Models investigating the evolutionary mechanisms that allow bacterial genes to be carried by plasmids suggest that plasmids should encode traits that are widely beneficial to many bacterial species, as for antibiotic resistance (Svara and Rankin, 2011; Lehtinen et al., 2021). While the fitness advantages conferred by antibiotic resistance genes are clear, the advantage of encoding genes on plasmids involved in broad categories of replication, recombination, repair, and cell motility require further investigation. Notably, COG functions for carbohydrate transport and metabolism (encompassing CAZymes), were found higher prevalence in the chromosome of *Curtobacterium* strains relative to their plasmids, however many glycoside hydrolases (important for degrading cellulose) were also identified on *Curtobacterium* plasmids. If particular CAZymes are beneficial within a particular ecotype, it may be that a plasmid-association could come under positive frequency-dependent selection, as has been modelled with antibiotic genes present on plasmids vs. chromosomes (Lehtinen et al., 2021).

Another type of trait that is often found on plasmids are those that confer pathogenicity or virulence (Johnson and Nolan, 2009). Interestingly, *Curtobacterium* plasmids from agricultural environments shared a high similarity in COG functional composition with those isolated from the desert and the Salton Sea environments. Initially, we expected the plasmids and

chromosomes from agricultural *Curtobacterium* (many of which are isolated from infected plants) would form a distinct ecotype. However, many of the agricultural strains group into the clade I along with strains from environments less impacted by humans. Recently, Chen et al., 2020 investigated three strains (MCBA15-007, MMLR14-002 and MMLR14-014) identified as *C. flaccumfaciens* that clustered with the plant pathogen strain P990 – a highly virulent bacteriocin-producing strain (Chen et al., 2020). Importantly, the P990 strain also had a plasmid that was a 147-kbp circular plasmid (pCff1) with 66.1% GC content as well as two circular plasmid-like DNA (sizes of 25 kb and 22 kb) found within the genome of the *C. flaccumfaciens* pv. *flaccumfacien* strain. Notably, some of the virulence/pathogenicity determinants that Chen et al., 2020 found on *Curtobacterium* strains associated with disease in plants were homologs of pathogenicity-determinant loci capable of producing 1,4-beta-xylanase (*xysA*), pectate lyase (*pelA1* and *pelA2*). Often these enzymes are thought important for carbohydrate utilization (i.e, CAZymes) associated with decomposition of senescent plant litter in soil communities (Chase et al., 2017; C. et al., 2022). We found that *Curtobacterium* plasmids from alpine, algae, and grassland environments had a relatively high prevalence of glycoside hydrolases, it could be that the type of CAZyme is indicative of the role of *Curtobacterium* as plant degrader versus pathogen. Additional comparative studies of *Curtobacterium* could identify plasmid content associated with habitat-specific adaptation such as the differences associated with commensal and pathogenic strains of *Escherichia* and host- versus free- living *Vibrio* species (Rasko et al., 2008; Chibani et al., 2020).

CONCLUSIONS

In sum, plasmids contribute to the variation in genomic diversity of *Curtobacterium*. Most plasmid research focuses on human-associated bacteria (such as those in the family

Enterobacteriaceae) and their resistance, virulence, or pathogenicity traits. However, we still lack a basic understanding about why other types of traits – such as those identified here – would be favored on plasmids in bacteria from free-living environments such as soil. Additional research in this and other well-studied free-living bacteria would help to broaden our understanding of the role of MGEs in bacterial evolution.

ACKNOWLEDGEMENTS

We thank Adam C. Martiny and Katrine L. Whiteson for helpful insights and comments on this work. This work was supported by a National Science Foundation Graduate Research Fellowship and UCI Chancellor's Club Fellowship to SSF, and by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research grants DE-SC0016410 and DE-SC0020382.

TABLES AND FIGURES

Table 3.1 PERMANOVA* testing the effects of environment and ecotype (clade designation) on *Curtobacterium* COG Functions by replicon type.

Trait/factors	DF	SS	MS	Pseudo-F	*P*-value	ECV (%)
COG - Plasmids						
Environment	6	158.91	26.485	1.7905	**0.017**	19.0
Ecotype	3	113.23	37.743	2.5515	**0.006**	37.2
Residuals	12	177.51	14.792			-
COG - Chromosomes						
Environment	9	202.65	22.517	2.7942	**0.001**	17.2
Ecotype	6	279.37	46.562	5.7781	**0.001**	38.2
Environment x Ecotype**	7	123.47	17.638	2.1888	**0.013**	20.0
Residuals	21	169.23	8.0584			-

*Significant *P*-values are boldface, indicates statistical significance with $P < 0.05$; *P*-values based on 999 permutations (lowest *P*-value possible 0.001); DF – degrees of freedom; SS – sum of squares; Pseudo-F – F value by permutation; ECV – estimated components of variation; Formula: Distance matrices ~ Environment + Ecotype.

**For this dataset there were enough levels of each factor to test for an Environment by Ecotype interaction.

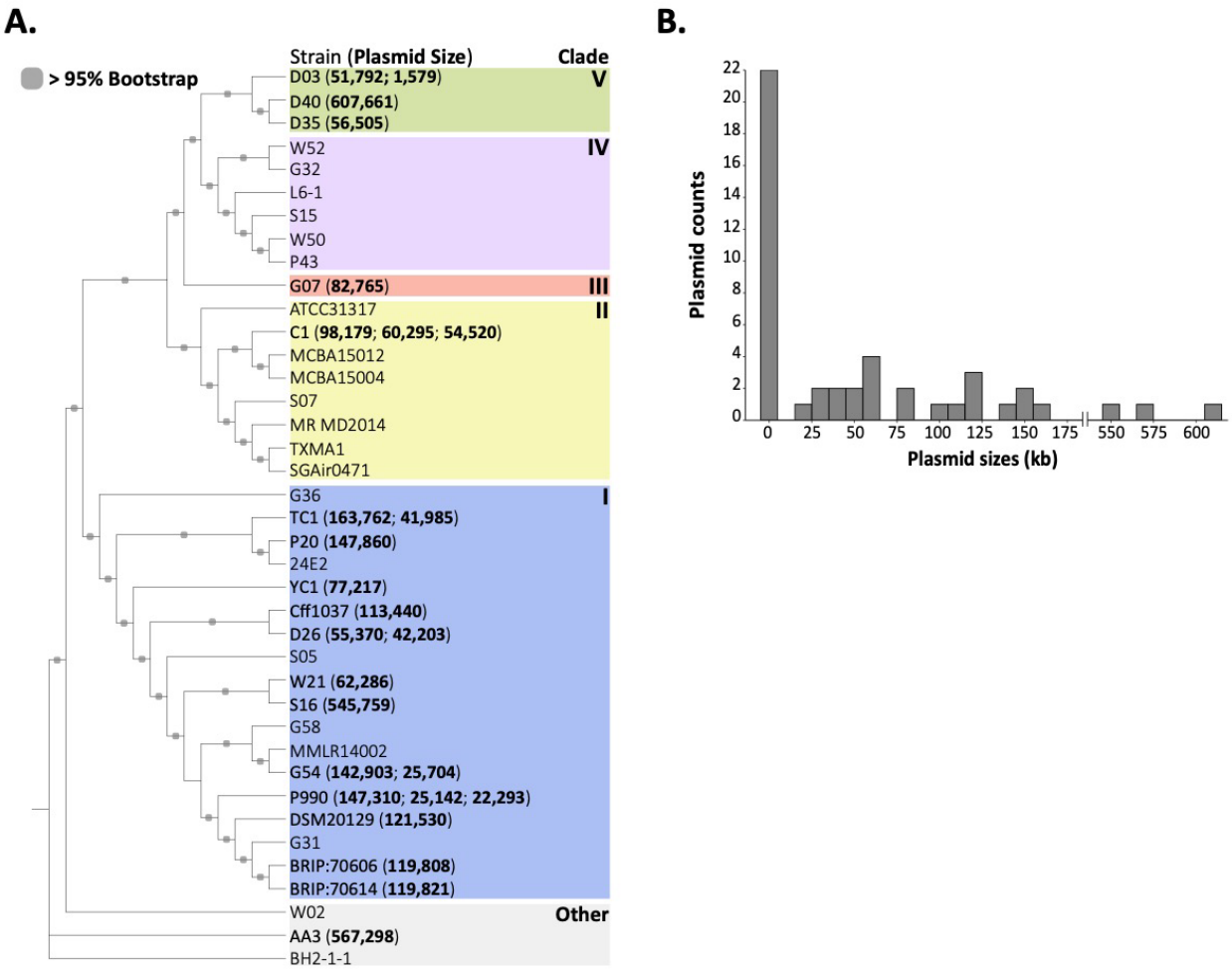

Figure 3.1 Cladogram (A) of complete chromosomes of *Curtobacterium* constructed from a phylogenomic analysis of 916 single-copy core genes. Bootstrap values are indicated by grey circles on branches. Bolded values next to strain identifiers are nucleotide lengths for circularized putative plasmid assemblies in bp. Ecotypes by clade designations are shaded in colors. Plasmid size (kb) distribution (B) across strains displayed in histogram.

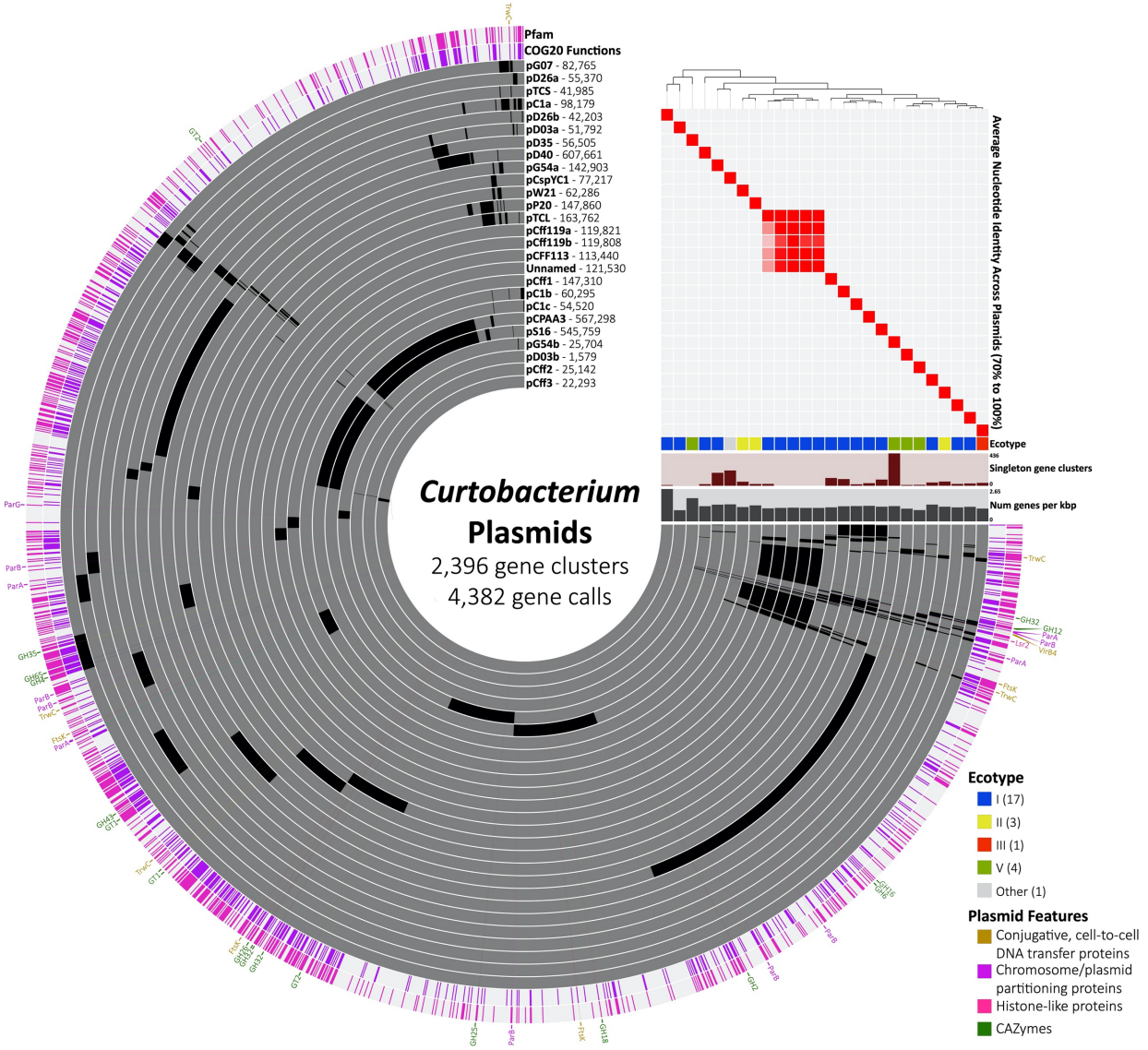

Figure 3.2 *Curtobacterium* plasmidome analysis. A visualization of 26 circularized/complete plasmids (rings) representing four host ecotypes and sizes ranging from approximately 1.5 – 607 kbp (nucleotide length in bp to right of plasmid ID). Common plasmid features for replication/separation during host cell division, carbohydrate utilization, and DNA protection are denoted by gene names around the outer ring. Pfam (pink) and COG20 functions (purple) that were identified for each gene call (not displayed) of a plasmid are displayed by colored bars. Gene clusters and singleton gene clusters (genes found only in one plasmid) are represented by black bars within gray rings. Gene clusters are organized by presence/absence and Euclidean distance using Ward linkage (see methods). The Average Nucleotide Identity (ANI) is normalized by the full sequence length for each plasmid to account large variations in plasmid sizes. The rings are ordered by normalized ANI measures as shown by the dendrogram above the ANI heatmap.

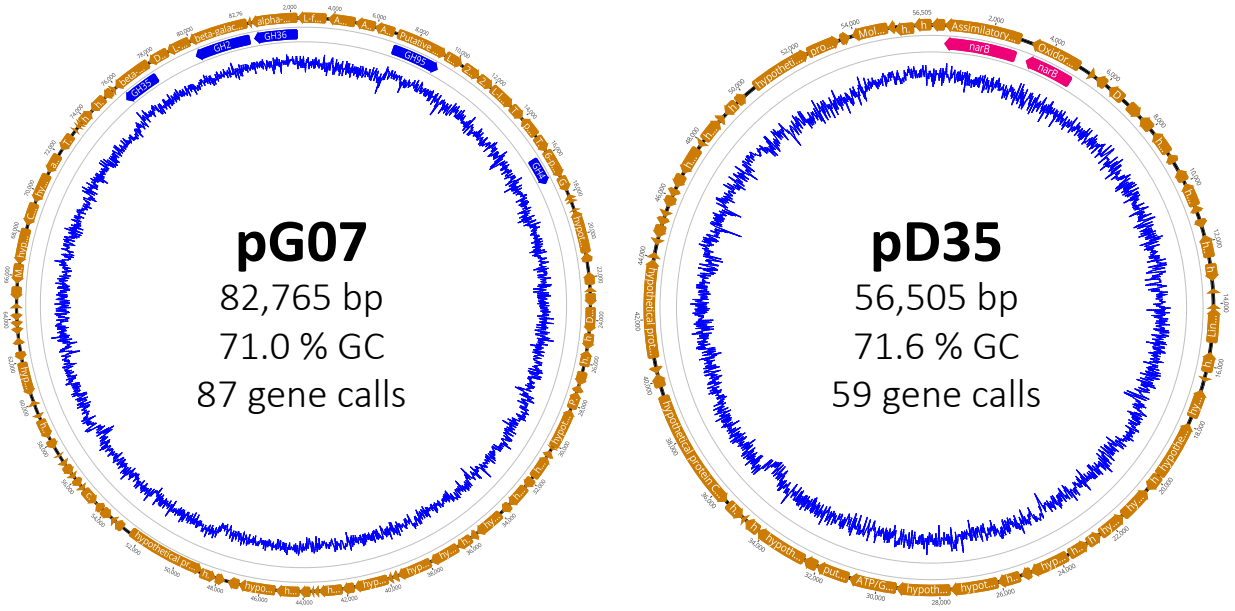

Figure 3.3 Schematic representation of two putative *Curtobacterium* plasmids assembled from a Grassland (left) and Desert (right) strains. The putative CAZymes identified in pG07 are glycoside hydrolases GH2,4,35,36,95 shown with blue arrows. The putative assimilatory nitrate reductases, encoded by *narB* are shown with pink arrows in the pD35.

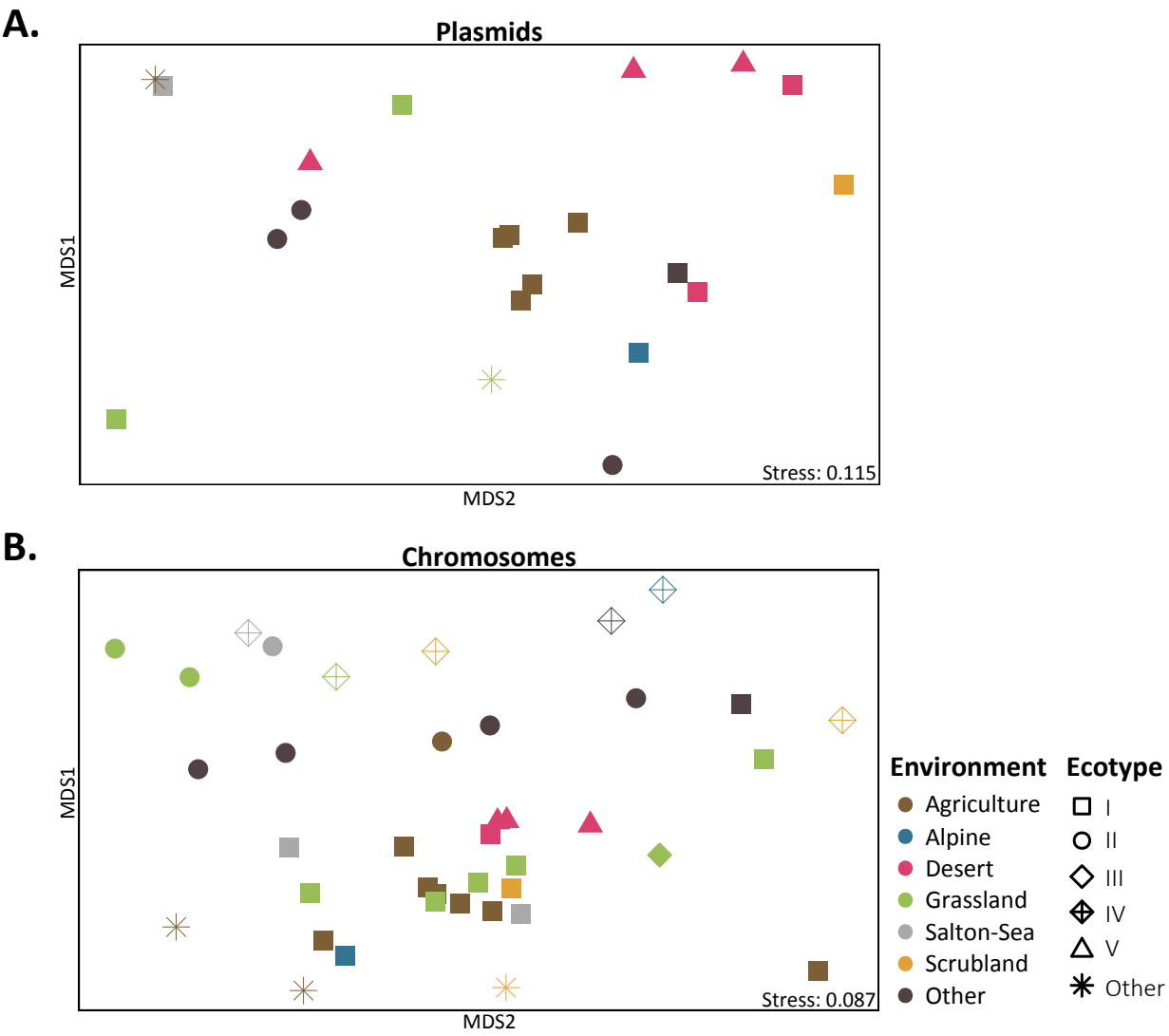

Figure 3.4 Non-metric multidimensional scaling (NMDS) ordinations of *Curtobacterium* complete sequence assemblies depicting (C) plasmid (D) chromosomal COG composition. Note –putative plasmids pCff2, pCff3, and pD03b are not shown since no COG functions were identified for their gene calls.

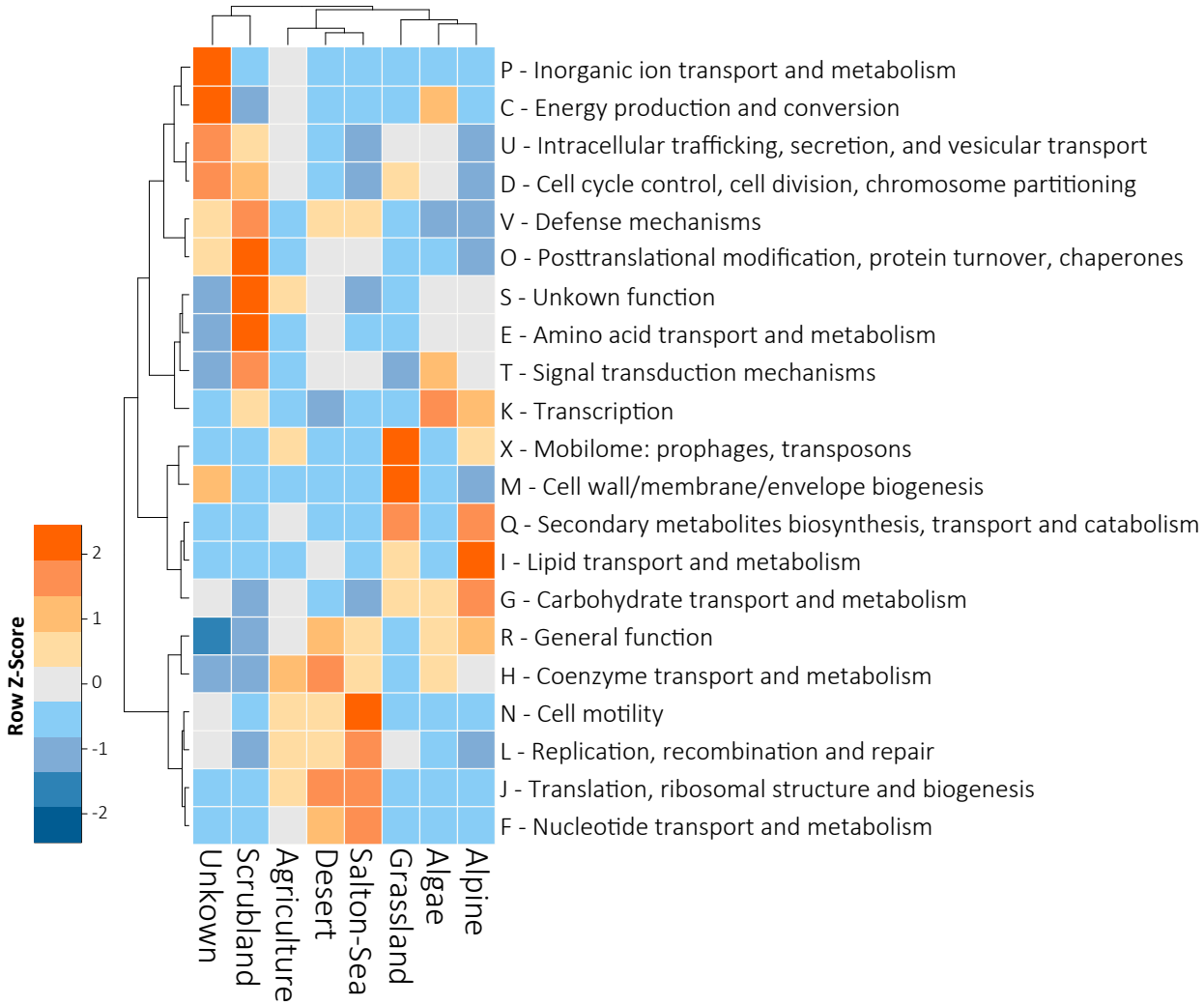

Figure 3.5 *Curtobacterium* Plasmid COG functions by environment. The normalized frequencies of COG functions by environment at broader COG category designations. To standardize for uneven plasmid sequences across environments, COG function counts were first converted into proportional abundances within an environment after removal of COG functions < 6 identified across all environments. COG abundance across environments were then normalized using Z-scores. COG categories with above mean (grey tiles) values are represented by tiles shaded in orange, while those with lower than mean values are shaded in blue. Log-Likelihood-Ratio test of independence for plasmid COG varying across environments was significant ($G(140) = 442.34$, $P < 0.001$). Note - plasmids pCff2, pCff3, and pD03b are not included in this analysis as no COG functions for these plasmids were identified.

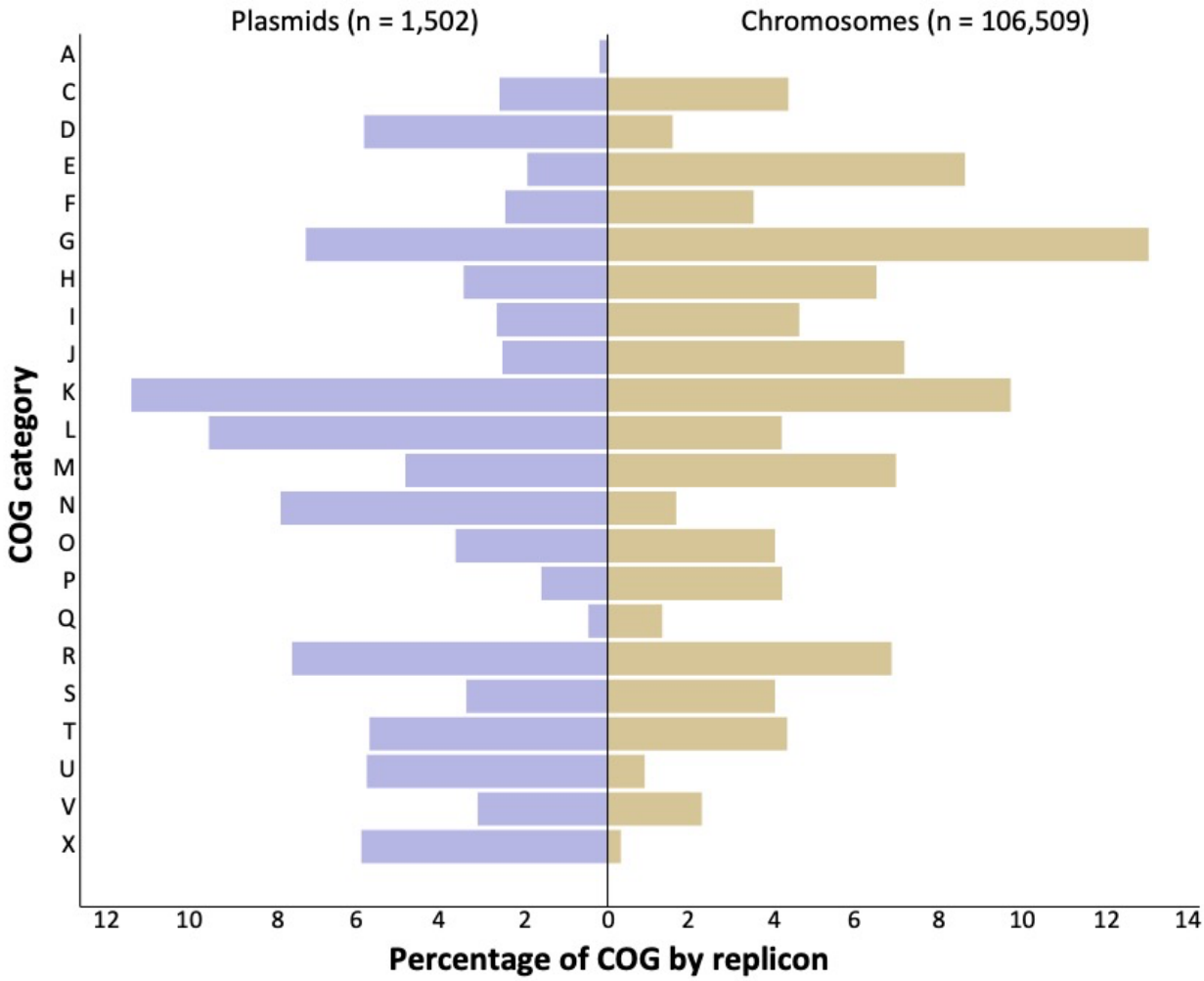

Figure 3.6 Overall COG category counts (Log_{10} scaled) for comparison across *Curtobacterium* replicon types. The total number of COG functions identified on *Curtobacterium* chromosomes and plasmids are shown in parentheses. Note – no COG functions for category Z were identified on *Curtobacterium* plasmids in this dataset, and plasmids pCff2, pCff3, and pD03b are not included in this analysis as no COG functions for these plasmids were identified.

REFERENCES

Acman M, van Dorp L, Santini JM, Balloux F. 2020. Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun* **11**(1): 2452. doi: 10.1038/s41467-020-16282-w

Adair EC, Parton WJ, Del Grosso SJ, Silver WL, Harmon ME, Hall SA, Burke IC, Hart SC. 2008. Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. *Glob Chang Biol*, in press. doi: 10.1111/j.1365-2486.2008.01674.x

Aerts R. 1997. Climate, Leaf Litter Chemistry and Leaf Litter Decomposition in Terrestrial Ecosystems: A Triangular Relationship. *Oikos* **79**(3): 439–449. [Nordic Society Oikos, Wiley]. doi: 10.2307/3546886

Allison SD. 2012. A trait-based approach for modelling microbial litter decomposition. *Ecol Lett* **15**(9): 1058–1070. John Wiley & Sons, Ltd. doi: https://doi.org/10.1111/j.1461-0248.2012.01807.x

Allison SD, Lu Y, Weihe C, Goulden ML, Martiny AC, Treseder KK, Martiny JBH. 2013. Microbial abundance and composition influence litter decomposition response to environmental change. *Ecology* **94**(3): 714–725. United States.

Allison SD, Martiny JBH. 2008. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* **105**(Supplement 1): 11512 LP – 11519. doi: 10.1073/pnas.0801925105

Alster CJ, German DP, Lu Y, Allison SD. 2013. Microbial enzymatic responses to drought and to nitrogen addition in a southern California grassland. *Soil Biol Biochem*, in press. doi: 10.1016/j.soilbio.2013.03.034

Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**(4): 775–808. American Society for Microbiology (ASM). doi: 10.1128/MMBR.00023-09

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**(1): 32–46. Wiley Online Library.

Anderson MJ. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, in press. doi: 10.1111/j.1541-0420.2005.00440.x

Anderson MJ. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online. doi: https://doi.org/10.1002/9781118445112.stat07841

Anderson MJ, Ellingsen KE, McArdle BH. 2006. Multivariate dispersion as a measure of beta diversity. *Ecol Lett*, in press. doi: 10.1111/j.1461-0248.2006.00926.x

Anderson MJ, Gorley RN, Clarke KR. 2008a. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. In: *Plymouth, UK*.

Anderson MJ, Gorley RN, Clarke KR. 2008b. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. In: *Plymouth, UK*.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**(W1): W16–W21. Oxford University Press. doi: 10.1093/nar/gkw387

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75. England. doi: 10.1186/1471-2164-9-75

Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F, Bürgmann H, Sørum H, Norström M, Pons M-N, et al. 2015. Tackling antibiotic resistance: the environmental framework. *Nat Rev Microbiol* **13**(5): 310–317. doi: 10.1038/nrmicro3439

Berg MP, Toby Kiers E, Driessen G, van der Heijden M, Kooi BW, Kuenen F, Liefting M, Verhoef HA, Ellers J. 2010. Adapt or disperse: Understanding species persistence in a changing world. *Glob Chang Biol*, in press. doi: 10.1111/j.1365-2486.2009.02014.x

Berlemont R, Allison SD, Weihe C, Lu Y, Brodie EL, Martiny JBH, Martiny AC. 2014. Cellulolytic potential under environmental changes in microbial communities from grassland litter . Frontiers in Microbiology .

Berlemont R, Martiny AC. 2013. Phylogenetic Distribution of Potential Cellulases in Bacteria. *Appl Environ Microbiol* **79**(5): 1545 LP – 1554. doi: 10.1128/AEM.03305-12

Berlemont R, Martiny AC. 2016. Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Comput Biol* **12**(12): e1005300. doi: 10.1371/journal.pcbi.1005300

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Caporaso JG. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**(1): 90. doi: 10.1186/s40168-018-0470-z

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Prepr* **6**: e27295v1. doi: 10.7287/peerj.preprints.27295v1

Bonham KS, Wolfe BE, Dutton RJ. 2017. Extensive horizontal gene transfer in cheese-associated bacteria. *Elife* **6**. England. doi: 10.7554/eLife.22144

Brito IL. 2021. Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* **19**(7): 442–453. doi: 10.1038/s41579-021-00534-7

Brockhurst MA, Harrison E. 2021. Ecological and evolutionary solutions to the plasmid paradox. *Trends Microbiol*, in press. doi: https://doi.org/10.1016/j.tim.2021.11.001

Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The Ecology and Evolution of Pangenomes. *Current Biology* **29**(20): R1094–R1103. doi: https://doi.org/10.1016/j.cub.2019.08.012

Brown Kav A, Rozov R, Bogumil D, Sørensen SJ, Hansen LH, Benhar I, Halperin E, Shamir R, Mizrahi I. 2020. Unravelling plasmidome distribution and interaction with its hosting microbiome. *Environ Microbiol* **22**(1): 32–44. John Wiley & Sons, Ltd. doi: https://doi.org/10.1111/1462-2920.14813

Brown Kav A, Sasson G, Jami E, Doron-Faigenboim A, Benhar I, Mizrahi I. 2012. Insights into the bovine rumen plasmidome. *Proc Natl Acad Sci U S A* **109**(14): 5452–5457. United States. doi: 10.1073/pnas.1116410109

Bruto M, James A, Petton B, Labreuche Y, Chenivesse S, Alunno-Bruscia M, Polz MF, Le Roux F. 2017. Vibrio crassostreae, a benign oyster colonizer turned into a pathogen after plasmid acquisition. *ISME J* **11**(4): 1043–1052. doi: 10.1038/ismej.2016.162

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**(1): 59–60. doi: 10.1038/nmeth.3176

C. SN, B. CA, S. FS, A. MA, C. W, D. AS, C. MA, H. MJB. 2022. Differential Response of Bacterial Microdiversity to Simulated Global Change. *Appl Environ Microbiol* **88**(6): e02429-21. American Society for Microbiology. doi: 10.1128/aem.02429-21

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**(7): 581. Nature Publishing Group. doi: 10.1038/nmeth.3869

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. [BMC Bioinformatics. 2009] - PubMed - NCBI. *BMC Bioinformatics*, in press.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**(5): 335. Nature Publishing Group. doi: 10.1038/nmeth.f.303

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**(8): 1621–1624. England. doi: 10.1038/ismej.2012.8

Carattoli A. 2009. Resistance plasmid families in Enterobacteriaceae. *Antimicrob Agents Chemother* **53**(6): 2227–2238. United States. doi: 10.1128/AAC.01707-08

Carattoli A. 2013. Plasmids and the spread of resistance. *Int J Med Microbiol* **303**(6–7): 298–304. Germany. doi: 10.1016/j.ijmm.2013.02.001

Castro HF, Classen AT, Austin EE, Norby RJ, Schadt CW. 2010. Soil Microbial Community Responses to Multiple Experimental Climate Change Drivers. *Appl Environ Microbiol* **76**(4): 999 LP – 1007. doi: 10.1128/AEM.02874-09

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, et al. 2019. Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol*, in press. doi: 10.1038/s41579-019-0222-5

Chapman SK, Newman GS. 2010. Biodiversity at the plant--soil interface: microbial abundance and community structure respond to litter mixing. *Oecologia* **162**(3): 763–769. doi: 10.1007/s00442-009-1498-3

Chase AB, Arevalo P, Polz MF, Berlemont R, Martiny JBH. 2016. Evidence for Ecological Flexibility in the Cosmopolitan Genus Curtobacterium . Frontiers in Microbiology .

Chase AB, Gomez-Lunar Z, Lopez AE, Li J, Allison SD, Martiny AC, Martiny JBH. 2018. Emergence of soil bacterial ecotypes along a climate gradient. *Environ Microbiol* **20**(11): 4112–4126. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/1462-2920.14405

Chase AB, Karaoz U, Brodie EL, Gomez-Lunar Z, Martiny AC, Martiny JBH. 2017. Microdiversity of an Abundant Terrestrial Bacterium Encompasses Extensive Variation in Ecologically Relevant Traits. *mBio* **8**(6). United States. doi: 10.1128/mBio.01809-17

Chase AB, Weihe C, Martiny JBH. 2021. Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. *Proceedings of the National Academy of Sciences* **118**(18): e2101254118. doi: 10.1073/pnas.2101254118

Chen G, Khojasteh M, Taheri-Dehkordi A, Taghavi SM, Rahimi T, Osdaghi E. 2020. Complete Genome Sequencing Provides Novel Insight Into the Virulence Repertories and Phylogenetic Position of Dry Beans Pathogen Curtobacterium flaccumfaciens pv. flaccumfaciens. *Phytopathology* **111**(2): 268–280. Scientific Societies. doi: 10.1094/PHYTO-06-20-0243-R

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17): i884–i890. doi: 10.1093/bioinformatics/bty560

Chibani CM, Roth O, Liesegang H, Wendling CC. 2020. Genomic variation among closely related Vibrio alginolyticus strains is located on mobile genetic elements. *BMC Genomics* **21**: 1–14. Springer.

Ciais P, Reichstein M, Viovy N, Granier A, Ogée J, Allard V, Aubinet M, Buchmann N, Bernhofer C, Carrara A, et al. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, in press. doi: 10.1038/nature03972

Cione NK, Padgett PE, Allen EB. 2002. Restoration of a Native Shrubland Impacted by Exotic Grasses, Frequent Fire, and Nitrogen Deposition in Southern California. *Restor Ecol* **10**(2): 376–384. John Wiley & Sons, Ltd (10.1111). doi: 10.1046/j.1526-100X.2002.02038.x

Clarke KR, Gorley RN. 2006a. PRIMER v6: Primer V6: User Manual/Tutorial.

Clarke KR, Gorley RN. 2006b. PRIMER v6: Primer V6: User Manual/Tutorial.

Classen AT, Sundqvist MK, Henning JA, Newman GS, Moore JAM, Cregger MA, Moorhead LC, Patterson CM. 2015. Direct and indirect effects of climate change on soil microbial and soil microbial-plant interactions: What lies ahead? *Ecosphere* **6**(8): art130. John Wiley & Sons, Ltd. doi: 10.1890/ES15-00217.1

Cleveland CC, Reed SC, Keller AB, Nemergut DR, O'Neill SP, Ostertag R, Vitousek PM. 2014. Litter quality versus soil microbial community controls over decomposition: A quantitative analysis. *Oecologia*, in press. doi: 10.1007/s00442-013-2758-9

Cohan FM. 2001. Bacterial species and speciation. *Syst Biol* **50**(4): 513–524. Society of Systematic Zoology.

Collard JM, Corbisier P, Diels L, Dong Q, Jeanthon C, Mergeay M, Taghavi S, van der Lelie D, Wilmotte A, Wuertz S. 1994. Plasmids for heavy metal resistance in Alcaligenes eutrophus CH34: mechanisms and applications. *FEMS Microbiol Rev* **14**(4): 405–414. England. doi: 10.1111/j.1574-6976.1994.tb00115.x

Conrad RE, Viver T, Gago JF, Hatt JK, Venter SN, Rossello-Mora R, Konstantinidis KT. 2022. Toward quantifying the adaptive role of bacterial pangenomes during environmental perturbations. *ISME J* **16**(5): 1222–1234. doi: 10.1038/s41396-021-01149-9

Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* **46**(W1): W246–W251. Oxford University Press. doi: 10.1093/nar/gky425

Craig S, Tyler A, Megan W, Austin H, Andres G, Matthew G, Mark P, David H, A. BP. 2022. F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal Escherichia coli. *mSphere* **5**(4): e00709-20. American Society for Microbiology. doi: 10.1128/mSphere.00709-20

Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun* **5**(1): 5471. doi: 10.1038/ncomms6471

Cruz-Martinez K, Suttle KB, Brodie EL, Power ME, Andersen GL, Banfield JF. 2009. Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *ISME J* **3**(6): 738–744. England. doi: 10.1038/ismej.2009.16

Curtiss III R. 1969. Bacterial conjugation. *Annu Rev Microbiol* **23**(1): 69–136. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Dai A. 2011. Drought under global warming: a review. *WIREs Climate Change* **2**(1): 45–65. John Wiley & Sons, Ltd. doi: 10.1002/wcc.81
Davison J. 1999. Genetic exchange between bacteria in the environment. Plasmid. doi: 10.1006/plas.1999.1421
Dewar AE, Thomas JL, Scott TW, Wild G, Griffin AS, West SA, Ghoul M. 2021. Plasmids do not consistently stabilize cooperation across bacteria but may promote broad pathogen host-range. *Nat Ecol Evol* **5**(12): 1624–1636. doi: 10.1038/s41559-021-01573-2
Dolph S, D. PT, R. GP. 1985. Ecological Character Displacement in Darwin's Finches. *Science (1979)* **227**(4690): 1056–1059. American Association for the Advancement of Science. doi: 10.1126/science.227.4690.1056
van Dongen S, Abreu-Goodger C. 2012. Using MCL to Extract Clusters from Networks BT - Bacterial Molecular Networks: Methods and Protocols. In: van Helden J, Toussaint A, Thieffry D, editors. New York, NY: Springer New York. p. 281–295. doi: 10.1007/978-1-61779-361-5_15
Doolittle WF, Papke RT. 2006. Genomics and the bacterial species problem. *Genome Biol* **7**(9): 1–7. BioMed Central.
Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res* **48**(D1): D561–D569. doi: 10.1093/nar/gkz1010
Dunivin TK, Choi J, Howe A, Shade A. 2019. RefSoil+: a Reference Database for Genes and Traits of Soil Plasmids. *mSystems*, in press. doi: 10.1128/msystems.00349-18
Eberhard WG. 1989. Why do bacterial plasmids carry some genes and not others? *Plasmid* **21**(3): 167–174. doi: https://doi.org/10.1016/0147-619X(89)90040-1
Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460–2461. England. doi: 10.1093/bioinformatics/btq461
Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol Lett*, in press. doi: 10.1111/j.1461-0248.2007.01113.x
Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**(1): 3–6. doi: 10.1038/s41564-020-00834-3
Evtushenko LI, Takeuchi M. 2006. The Family Microbacteriaceae. In: *The Prokaryotes*. doi: 10.1007/0-387-30743-5_43
Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**(19): 3047–3048. doi: 10.1093/bioinformatics/btw354
Fahey C, Koyama A, Antunes PM, Dunfield K, Flory SL. 2020. Plant communities mediate the interactive effects of invasion and drought on soil microbial communities. *ISME Journal*, in press. doi: 10.1038/s41396-020-0614-6
Farshad S, Ranjbar R, Japoni A, Hosseini M, Anvarinejad M, Mohammadzadegan R. 2012. Microbial susceptibility, virulence factors, and plasmid profiles of uropathogenic Escherichia coli strains isolated from children in Jahrom, Iran. *Arch Iran Med*, in press.
Fatoba DO, Amoako DG, Akebe ALK, Ismail A, Essack SY. 2022. Genomic analysis of antibiotic-resistant Enterococcus spp. reveals novel enterococci strains and the spread of

plasmid-borne Tet(M), Tet(L) and Erm(B) genes from chicken litter to agricultural soil in South Africa. *J Environ Manage* **302**: 114101. doi: https://doi.org/10.1016/j.jenvman.2021.114101

Fernández-Alonso MJ, Curiel Yuste J, Kitzler B, Ortiz C, Rubio A. 2018. Changes in litter chemistry associated with global change-driven forest succession resulted in time-decoupled responses of soil carbon and nitrogen cycles. *Soil Biol Biochem* **120**: 200–211. doi: https://doi.org/10.1016/j.soilbio.2018.02.013

Finks SS, Weihe C, Kimball S, Allison SD, Martiny AC, Treseder KK, Martiny JBH. 2021. Microbial community response to a decade of simulated global changes depends on the plant community. *Elementa: Science of the Anthropocene* **9**(1). doi: 10.1525/elementa.2021.00124

Fischer DG, Chapman SK, Classen AT, Gehring CA, Grady KC, Schweitzer JA, Whitham TG. 2014. Plant genetic effects on soils under climate change. *Plant Soil* **379**(1): 1–19. doi: 10.1007/s11104-013-1972-x

Flemming H-C, Wuertz S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* **17**(4): 247–260. doi: 10.1038/s41579-019-0158-9

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**(9): 722–732. doi: 10.1038/nrmicro1235

Fuchslueger L, Bahn M, Fritz K, Hasibeder R, Richter A. 2014. Experimental drought reduces the transfer of recently fixed plant carbon to soil microbes and alters the bacterial community composition in a mountain meadow. *New Phytologist* **201**(3): 916–927. John Wiley & Sons, Ltd. doi: 10.1111/nph.12569

Fulthorpe RR, Wyndham RC. 1991. Transfer and Expression of the Catabolic Plasmid pBRC60 in Wild Bacterial Recipients in a Freshwater Ecosystem. *Appl Environ Microbiol* **57**(5): 1546–1553. doi: 10.1128/AEM.57.5.1546-1553.1991

Gaertner M, Den Breeyen A, Hui C, Richardson DM. 2009. Impacts of alien plant invasions on species richness in Mediterranean-type ecosystems: a meta-analysis. *Progress in Physical Geography: Earth and Environment* **33**(3): 319–338. SAGE Publications Ltd. doi: 10.1177/0309133309341607

Galata V, Fehlmann T, Backes C, Keller A. 2019. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* **47**(D1): D195–D202. doi: 10.1093/nar/gky1050

Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin E V. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* **49**(D1): D274–D281. doi: 10.1093/nar/gkaa1018

Gao YZ, Chen Q, Lin S, Giese M, Brueck H. 2011. Resource manipulation effects on net primary production, biomass allocation and rain-use efficiency of two semiarid grassland sites in Inner Mongolia, China. *Oecologia*, in press. doi: 10.1007/s00442-010-1890-z

Garcillán-Barcia MP, Francia MV, de La Cruz F. 2009. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* **33**(3): 657–687. doi: 10.1111/j.1574-6976.2009.00168.x

Garcillán-Barcia MP, Redondo-Salvo S, Vielva L, de la Cruz F. 2020. MOBscan: Automated Annotation of MOB Relaxases. *Methods Mol Biol* **2075**: 295–308. United States. doi: 10.1007/978-1-4939-9877-7_21

Ghaly TM, Chow L, Asher AJ, Waldron LS, Gillings MR. 2017. Evolution of class 1 integrons: Mobilization and dispersal via food-borne bacteria. *PLoS One* **12**(6): e0179169. Public Library of Science.

Giraldo R. 2003. Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev* **26**(5): 533–554. doi: 10.1111/j.1574-6976.2003.tb00629.x

Glassman SI, Martiny JBH. 2018. Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. Tringe SG, editor. *mSphere* **3**(4): e00148-18. doi: 10.1128/mSphere.00148-18

Glassman SI, Weihe C, Li J, Albright MBN, Looby CI, Martiny AC, Treseder KK, Allison SD, Martiny JBH. 2018. Decomposition responses to climate depend on microbial community composition. *Proceedings of the National Academy of Sciences* **115**(47): 11994 LP – 11999. doi: 10.1073/pnas.1811269115

Gleixner G. 2013. Soil organic matter dynamics: a biological perspective derived from the use of compound-specific isotopes studies. *Ecol Res* **28**(5): 683–695. Springer.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**(1): 81–91. Microbiology Society. doi: 10.1099/ijs.0.64483-0

Graff P, Gundel PE, Salvat A, Cristos D, Chaneton EJ. 2020. Protection offered by leaf fungal endophytes to an invasive species against native herbivores depends on soil nutrients. *Journal of Ecology*, in press. doi: 10.1111/1365-2745.13371

Griffin D, Anchukaitis KJ. 2014. How unusual is the 2012-2014 California drought? *Geophys Res Lett*, in press. doi: 10.1002/2014GL062433

Griffin-Nolan RJ, Blumenthal DM, Collins SL, Farkas TE, Hoffman AM, Mueller KE, Ocheltree TW, Smith MD, Whitney KD, Knapp AK. 2019. Shifts in plant functional composition following long-term drought in grasslands. Jones H, editor. *Journal of Ecology* **107**(5): 2133–2148. doi: 10.1111/1365-2745.13252

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**(8): 1072–1075. doi: 10.1093/bioinformatics/btt086

Gutknecht JLM, Field CB, Balser TC. 2012. Microbial communities and their responses to simulated global change fluctuate greatly over multiple years. *Glob Chang Biol* **18**(7): 2256–2269. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/j.1365-2486.2012.02686.x

H. MJB, E. JS, T. LJ, C. MA. 2015. Microbiomes in light of traits: A phylogenetic perspective. *Science (1979)* **350**(6261): aac9323. American Association for the Advancement of Science. doi: 10.1126/science.aac9323

Hairston Jr NG, Ellner SP, Geber MA, Yoshida T, Fox JA. 2005. Rapid evolution and the convergence of ecological and evolutionary time. *Ecol Lett* **8**(10): 1114–1127. John Wiley & Sons, Ltd. doi: https://doi.org/10.1111/j.1461-0248.2005.00812.x

Hall JPJ, Botelho J, Cazares A, Baltrus DA. 2022. What makes a megaplasmid? *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**(1842): 20200472. Royal Society. doi: 10.1098/rstb.2020.0472

Hall JPJ, Brockhurst MA, Harrison E. 2017. Sampling the mobile gene pool: Innovation via horizontal gene transfer in bacteria. Philosophical Transactions of the Royal Society B: Biological Sciences. doi: 10.1098/rstb.2016.0424

Hall JPJ, Wright RCT, Harrison E, Muddiman KJ, Wood AJ, Paterson S, Brockhurst MA. 2021. Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. *PLoS Biol* **19**(10): e3001225. Public Library of Science.

Hanka B, Marlene H, Thomas K, Mohammed W, Volker S, A. BA, Boyke B, Cathrin S, Jörg O, Erika K, et al. 2022. Extremophile Metal Resistance: Plasmid-Encoded Functions in

Streptomyces mirabilis. *Appl Environ Microbiol* **88**(11): e00085-22. American Society for Microbiology. doi: 10.1128/aem.00085-22
Harrison E, Brockhurst MA. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol* **20**(6): 262–267. England. doi: 10.1016/j.tim.2012.04.003
Heuer H, Smalla K. 2012. Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol Rev* **36**(6): 1083–1104. England. doi: 10.1111/j.1574-6976.2012.00337.x
Le Houérou HN. 1996. Climate change, drought and desertification. *J Arid Environ* **34**(2): 133–185. doi: https://doi.org/10.1006/jare.1996.0099
Hu F, O'Hara JA, Rivera JI, Doi Y. 2014. Molecular Features of Community-Associated Extended-Spectrum-β-Lactamase-Producing Escherichia coli Strains in the United States. *Antimicrob Agents Chemother* **58**(11): 6953–6957. doi: 10.1128/AAC.03321-14
Hu Y, Yang X, Li J, Lv N, Liu F, Wu J, Lin IYC, Wu N, Weimer BC, Gao GF, et al. 2016. The bacterial mobile resistome transfer network connecting the animal and human microbiomes. *Appl Environ Microbiol*, in press. doi: 10.1128/AEM.01802-16
Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**(1): 119. doi: 10.1186/1471-2105-11-119
I. CB, R. AT, E. SJ. 2022. Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci Adv* **1**(1): e1400082. American Association for the Advancement of Science. doi: 10.1126/sciadv.1400082
IPCC 2014. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. Geneva, Switzerland.
J. BC, Diya S, Hirokazu Y, L. BM, M. RL, A. V der AG, M. TE. 2013. Diverse Broad-Host-Range Plasmids from Freshwater Carry Few Accessory Genes. *Appl Environ Microbiol* **79**(24): 7684–7695. American Society for Microbiology. doi: 10.1128/AEM.02252-13
J. HC, M. HR, Melanie B. 2021. IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* **1**(2): e00038-16. American Society for Microbiology. doi: 10.1128/mSphere.00038-16
J. NR, L. MS, K. DD, H. VJ, G. MH, Murat EA, L. SM, B. HG. 2015. Sewage Reflects the Microbiomes of Human Populations. *mBio* **6**(2): e02574-14. American Society for Microbiology. doi: 10.1128/mBio.02574-14
J. WM, Noah R, E. LR. 2013. Long-Term Dynamics of Adaptation in Asexual Populations. *Science (1979)* **342**(6164): 1364–1367. American Association for the Advancement of Science. doi: 10.1126/science.1243357
Jackson R, McCawley W. 1997. The First Angelinos: The Gabrielino Indians of Los Angeles. *Ethnohistory* **44**(2): 415. doi: 10.2307/483386
Jacoby GA, Sutton L. 1991. Properties of plasmids responsible for production of extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* **35**(1): 164 LP – 169. doi: 10.1128/AAC.35.1.164
Johnsborg O, Eldholm V, Håvarstein LS. 2007. Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**(10): 767–778. doi: https://doi.org/10.1016/j.resmic.2007.09.004
Johnson TJ, Nolan LK. 2009. Pathogenomics of the Virulence Plasmids of Escherichia coli. *Microbiology and Molecular Biology Reviews*, in press. doi: 10.1128/mmbr.00015-09

Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK. 2007. Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates. *Appl Environ Microbiol* **73**(6): 1976–1983. United States. doi: 10.1128/AEM.02171-06

Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **12**(3): 181–196. Nature Publishing Group.

Kara Ö, Bolat I, Çakiroğlu K, Öztürk M. 2008. Plant canopy effects on litter accumulation and soil microbial biomass in two temperate forests. *Biol Fertil Soils*, in press. doi: 10.1007/s00374-008-0327-x

Kimball S, Funk JL, Spasojevic MJ, Suding KN, Parker S, Goulden ML. 2016. Can functional traits predict plant community response to global change? *Ecosphere* **7**(12): e01602. John Wiley & Sons, Ltd. doi: 10.1002/ecs2.1602

Kimball S, Goulden ML, Suding KN, Parker S. 2014. Altered water and nitrogen input shifts succession in a southern California coastal sage community. *Ecol Appl* **24**(6): 1390–1404. United States.

Kinugasa T, Tsunekawa A, Shinoda M. 2012. Increasing nitrogen deposition enhances post-drought recovery of grassland productivity in the Mongolian steppe. *Oecologia* **170**(3): 857–865. doi: 10.1007/s00442-012-2354-4

Kirchberger PC, Schmidt ML, Ochman H. 2020. The Ingenuity of Bacterial Genomes. *Annu Rev Microbiol* **74**(1): 815–834. Annual Reviews. doi: 10.1146/annurev-micro-020518-115822

Klümper U, Riber L, Dechesne A, Sannazzarro A, Hansen LH, Sørensen SJ, Smets BF. 2015. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J* **9**(4): 934–945. doi: 10.1038/ismej.2014.191

Kominoski JS, Hoellein TJ, Kelly JJ, Pringle CM. 2009. Does mixing litter of different qualities alter stream microbial diversity and functioning on individual litter species? *Oikos* **118**(3): 457–463. doi: 10.1111/j.1600-0706.2008.17222.x

Krause S, Le Roux X, Niklaus PA, Van Bodegom PM, Lennon JT, Bertilsson S, Grossart H-P, Philippot L, Bodelier PLE. 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning . Frontiers in Microbiology .

Kuypers MMM, Marchant HK, Kartal B. 2018. The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**(5): 263–276. doi: 10.1038/nrmicro.2018.9

Land ML, Hyatt D, Jun S-R, Kora GH, Hauser LJ, Lukjancenko O, Ussery DW. 2014. Quality scores for 32,000 genomes. *Stand Genomic Sci* **9**(1): 20. doi: 10.1186/1944-3277-9-20

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**(20): 6955–6959. doi: 10.1073/pnas.82.20.6955

Lau JA, Lennon JT. 2012. Rapid responses of soil microorganisms improve plant fitness in novel environments. *Proc Natl Acad Sci U S A*, in press. doi: 10.1073/pnas.1202319109

Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, Barraclough TG. 2012. Species interactions alter evolutionary responses to a novel environment. *PLoS Biol* **10**(5): e1001330–e1001330. Public Library of Science. doi: 10.1371/journal.pbio.1001330

LeBauer DS, Treseder KK. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, in press. doi: 10.1890/06-2057.1

Legay N, Lavorel S, Baxendale C, Krainer U, Bahn M, Binet M-N, Cantarel AAM, Colace M-P, Foulquier A, Kastl E-M, et al. 2016. Influence of plant traits, soil microbial properties, and

abiotic parameters on nitrogen turnover of grassland ecosystems. *Ecosphere* **7**(11): e01448. John Wiley & Sons, Ltd. doi: 10.1002/ecs2.1448
Lehtinen S, Huisman JS, Bonhoeffer S. 2021. Evolutionary mechanisms that determine which bacterial genes are carried on plasmids. *Evol Lett* **5**(3): 290–301. John Wiley and Sons Inc. doi: 10.1002/evl3.226
Lerner A, Matthias T, Aminov R. 2017. Potential Effects of Horizontal Gene Exchange in the Human Gut . Frontiers in Immunology .
Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**(W1): W293–W296. doi: 10.1093/nar/gkab301
Li J-J, Spychala CN, Hu F, Sheng J-F, Doi Y. 2015. Complete Nucleotide Sequences of bla CTX-M -Harboring IncF Plasmids from Community-Associated Escherichia coli Strains in the United States. *Antimicrob Agents Chemother* **59**(6): 3002–3007. doi: 10.1128/AAC.04772-14
Li L, Dechesne A, Madsen JS, Nesme J, Sørensen SJ, Smets BF. 2020. Plasmids persist in a microbial community by providing fitness benefit to multiple phylotypes. *ISME Journal*, in press. doi: 10.1038/s41396-020-0596-4
Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**(5950): 289–293. doi: 10.1126/science.1181369
Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou H-Y. 2019. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* **47**(D1): D660–D665. doi: 10.1093/nar/gky1123
Long SR. 1996. Rhizobium symbiosis: nod factors in perspective. *Plant Cell* **8**(10): 1885–1898. doi: 10.1105/tpc.8.10.1885
Looby CI, Maltz MR, Treseder KK. 2016. Belowground responses to elevation in a changing cloud forest. *Ecol Evol* **6**(7): 1996–2009. doi: 10.1002/ece3.2025
Luo W, Xu Z, Riber L, Hansen LH, Sørensen SJ. 2016. Diverse gene functions in a soil mobilome. *Soil Biol Biochem*, in press. doi: 10.1016/j.soilbio.2016.07.018
M. ABM, Mimi T, Fred H. 1999. The Virulence Plasmid of Salmonella typhimurium Is Self-Transmissible. *J Bacteriol* **181**(4): 1364–1368. American Society for Microbiology. doi: 10.1128/JB.181.4.1364-1368.1999
Malik Ashish A, Martiny JBH, Brodie EL, Martiny AC, Treseder KK, Allison SD. 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J* **14**(1): 1–9. doi: 10.1038/s41396-019-0510-0
Malik Ashish A., Swenson T, Weihe C, Morrison EW, Martiny JBH, Brodie EL, Northen TR, Allison SD. 2020. Drought and plant litter chemistry alter microbial gene expression and metabolite production. *ISME J* **14**(9): 2236–2247. doi: 10.1038/s41396-020-0683-6
Martiny JB, Martiny AC, Weihe C, Lu Y, Berlemont R, Brodie EL, Goulden ML, Treseder KK, Allison SD. 2017. Microbial legacies alter decomposition in response to simulated global change. *ISME J* **11**(2): 490–499. doi: 10.1038/ismej.2016.122
Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol* **17**(10): 458–466. doi: https://doi.org/10.1016/j.tim.2009.07.004

Matulich KL, Martiny JBH. 2015. Microbial composition alters the response of litter decomposition to environmental change. *Ecology* **96**(1): 154–163. John Wiley & Sons, Ltd. doi: 10.1890/14-0357.1

Matulich KL, Weihe C, Allison SD, Amend AS, Berlemont R, Goulden ML, Kimball S, Martiny AC, Martiny JBH. 2015. Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. *ISME J* **9**(11): 2477–2489. doi: 10.1038/ismej.2015.58

McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal*, in press. doi: 10.1038/ismej.2011.139

Meng M, Li Y, Yao H. 2022. Plasmid-Mediated Transfer of Antibiotic Resistance Genes in Soil. Antibiotics . doi: 10.3390/antibiotics11040525

Michael R, Ramon R-M. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* **106**(45): 19126–19131. Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.0906412106

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**(5): 1530–1534. doi: 10.1093/molbev/msaa015

Mishra AK, Singh VP. 2010. A review of drought concepts. *J Hydrol (Amst)* **391**(1): 202–216. doi: https://doi.org/10.1016/j.jhydrol.2010.07.012

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**(D1): D412–D419. doi: 10.1093/nar/gkaa913

Mooshammer M, Wanek W, Hämmerle I, Fuchslueger L, Hofhansl F, Knoltsch A, Schnecker J, Takriti M, Watzka M, Wild B, et al. 2014. Adjustment of microbial nitrogen use efficiency to carbon:nitrogen imbalances regulates soil nitrogen cycling. *Nat Commun* **5**(1): 3694. doi: 10.1038/ncomms4694

Moretto JAS, Braz VS, Furlan JPR, Stehling EG. 2019. Plasmids associated with heavy metal resistance and herbicide degradation potential in bacterial isolates obtained from two Brazilian regions. *Environ Monit Assess* **191**(5). Springer International Publishing. doi: 10.1007/s10661-019-7461-9

Nakatsu CH, Fulthorpe RR, Holland BA, Peel MC, Wyndham RC. 1995. The phylogenetic distribution of a transposable dioxygenase from the Niagara River watershed. *Mol Ecol* **4**(5): 593–603. England. doi: 10.1111/j.1365-294x.1995.tb00259.x

Nelson MB, Martiny AC, Martiny JBH. 2016. Global biogeography of microbial nitrogen-cycling traits in soil. *Proc Natl Acad Sci U S A*, in press. doi: 10.1073/pnas.1601070113

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**(1): 268–274. doi: 10.1093/molbev/msu300

Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, et al. 2018. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* **47**(D1): D259–D264. doi: 10.1093/nar/gky1022

Nordmann P, Naas T, Poirel L. 2011. Global spread of Carbapenemase-producing Enterobacteriaceae. *Emerg Infect Dis* **17**(10): 1791–1798. United States. doi: 10.3201/eid1710.110655

Norman A, Hansen LH, Sørensen SJ. 2009. Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1527): 2275–2289. Royal Society. doi: 10.1098/rstb.2009.0037

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2019. vegan: Community Ecology Package. R package version 2.5-2. *Cran R*, in press.

Ostle NJ, Smith P, Fisher R, Ian Woodward F, Fisher JB, Smith JU, Galbraith D, Levy P, Meir P, McNamara NP, et al. 2009. Integrating plant–soil interactions into global carbon cycle models. *Journal of Ecology* **97**(5): 851–863. doi: doi:10.1111/j.1365-2745.2009.01547.x

Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* **18**(5): 1403–1414. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/1462-2920.13023

Parmesan C. 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annu Rev Ecol Evol Syst* **37**(1): 637–669. Annual Reviews. doi: 10.1146/annurev.ecolsys.37.091305.110100

Partridge SR, Kwong SM, Firth N, Jensen SO. 2018. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* **31**(4): e00088-17. American Society for Microbiology. doi: 10.1128/CMR.00088-17

Perez MF, Kurth D, Farías ME, Soria MN, Castillo Villamizar GA, Poehlein A, Daniel R, Dib JR. 2020. First Report on the Plasmidome From a High-Altitude Lake of the Andean Puna . Frontiers in Microbiology .

Philippon A, Arlet G, Lagrange PH. 1994. Origin and impact of plasmid-mediated extended-spectrum beta-lactamases. *European Journal of Clinical Microbiology and Infectious Diseases* **13**(1): S17–S29. doi: 10.1007/BF02390681

Pinto UM, Pappas KM, Winans SC. 2012. The ABCs of plasmid replication and segregation. *Nat Rev Microbiol* **10**(11): 755–765. doi: 10.1038/nrmicro2882

Potts DL, Suding KN, Winston GC, Rocha AV, Goulden ML. 2012. Ecological effects of experimental drought and prescribed fire in a southern California coastal grassland. *J Arid Environ* **81**: 59–66. Academic Press. doi: 10.1016/J.JARIDENV.2012.01.007

Preece C, Verbruggen E, Liu L, Weedon JT, Peñuelas J. 2019. Effects of past and current drought on the composition and diversity of soil microbial communities. *Soil Biol Biochem*, in press. doi: 10.1016/j.soilbio.2018.12.022

Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8**(1): 12–24. The Royal Society of Chemistry. doi: 10.1039/C5AY02550H

Quistad SD, Doulcier G, Rainey PB. 2020. Experimental manipulation of selfish genetic elements links genes to microbial community function. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**(1798): 20190681. Royal Society. doi: 10.1098/rstb.2019.0681

R. 2018. R core team (2018). *R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL http://www R-project org*, in press.

R Core Team (2020). 2020. R: A language and environment for statistical computing. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rahube TO, Viana LS, Koraimann G, Yost CK. 2014. Characterization and comparative analysis of antibiotic resistance plasmids isolated from a wastewater treatment plant . Frontiers in Microbiology .
Rahube TO, Yost CK. 2010. Antibiotic resistance plasmids in wastewater treatment plants and their possible dissemination into the environment. *Afr J Biotechnol* **9**(54): 9183–9190.
Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* **394**(6688): 69–72. Nature Publishing Group.
Rankin DJ, Rocha EPC, Brown SP. 2011. What traits are carried on mobile genetic elements, and why. Heredity. doi: 10.1038/hdy.2010.24
Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, et al. 2008. The pangenome structure of Escherichia coli: Comparative genomic analysis of E. coli commensal and pathogenic isolates. *J Bacteriol*, in press. doi: 10.1128/JB.00619-08
Redondo-Salvo S, Fernández-López R, Ruiz R, Vielva L, de Toro M, Rocha EPC, Garcillán-Barcia MP, de la Cruz F. 2020. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat Commun* **11**(1): 3602. doi: 10.1038/s41467-020-17278-2
Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D. 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Science of The Total Environment* **447**: 345–360. doi: https://doi.org/10.1016/j.scitotenv.2013.01.032
Rodriguez RJ, White JF, Arnold AE, Redman RS. 2009. Fungal endophytes: Diversity and functional roles: Tansley review. New Phytologist. doi: 10.1111/j.1469-8137.2009.02773.x
Romaniuk K, Golec P, Dziewit L. 2018. Insight Into the Diversity and Possible Role of Plasmids in the Adaptation of Psychrotolerant and Metalotolerant Arthrobacter spp. to Extreme Antarctic Environments. *Front Microbiol*, in press. doi: 10.3389/fmicb.2018.03144
Rosenzweig C, Karoly D, Vicarelli M, Neofotis P, Wu Q, Casassa G, Menzel A, Root TL, Estrella N, Seguin B, et al. 2008. Attributing physical and biological impacts to anthropogenic climate change. *Nature* **453**(7193): 353–357. England. doi: 10.1038/nature06937
Rotger R, Casadesús J. 1999. The virulence plasmids of Salmonella. *Int Microbiol* **2**(3): 177–184. Switzerland.
Rouifed S, Handa IT, David J-F, Hättenschwiler S. 2010. The importance of biotic factors in predicting global change effects on decomposition of temperate forest leaf litter. *Oecologia* **163**(1): 247–256. doi: 10.1007/s00442-009-1528-1
Ryo M, Yukari S, Michihiro I, Yoshiyuki O, Yuji N, Masataka T. 2006. Complete Nucleotide Sequence of an Exogenously Isolated Plasmid, pLB1, Involved in γ-Hexachlorocyclohexane Degradation. *Appl Environ Microbiol* **72**(11): 6923–6933. American Society for Microbiology. doi: 10.1128/AEM.01531-06
Salazar G. 2020. EcolUtils: Utilities for community ecology analysis.
San Millan A. 2018. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends Microbiol* **26**(12): 978–985. doi: https://doi.org/10.1016/j.tim.2018.06.007

Santonja M, Rancon A, Fromin N, Baldy V, Hättenschwiler S, Fernandez C, Montès N, Mirleau P. 2017. Plant litter diversity increases microbial abundance, fungal diversity, and carbon and nitrogen cycling in a Mediterranean shrubland. *Soil Biol Biochem* **111**: 124–134. doi: https://doi.org/10.1016/j.soilbio.2017.04.006

Sayer EJ, Oliver AE, Fridley JD, Askew AP, Mills RTE, Grime JP. 2017. Links between soil microbial communities and plant traits in a species-rich grassland under long-term climate change. *Ecol Evol* **7**(3): 855–862. John Wiley & Sons, Ltd. doi: 10.1002/ece3.2700

Sayler GS, Hooper SW, Layton AC, King JMH. 1990. Catabolic plasmids of environmental and ecological significance. Microbial Ecology. doi: 10.1007/BF02015050

Schimel JP, Schaeffer SM. 2012. Microbial control over carbon cycling in soil. *Front Microbiol* **3**: 348. Frontiers Research Foundation. doi: 10.3389/fmicb.2012.00348

Schmidt PA, Schmitt I, Otte J, Bandow C, Römbke J, Bálint M, Rolshausen G. 2018. Season-Long Experimental Drought Alters Fungal Community Composition but Not Diversity in a Grassland Soil. *Microb Ecol*, in press. doi: 10.1007/s00248-017-1047-2

Schneider CL. 2021. Bacteriophage-mediated horizontal gene transfer: transduction. *Bacteriophages: Biology, Technology, Therapy*: 151–192. Springer.

Sen D, Van der Auwera GA, Rogers LM, Thomas CM, Brown CJ, Top EM. 2011. Broad-host-range plasmids from agricultural soils have IncP-1 backbones with diverse accessory genes. *Appl Environ Microbiol* **77**(22): 7975–7983. American Society for Microbiology. doi: 10.1128/AEM.05439-11

Sentchilo V, Mayer AP, Guy L, Miyazaki R, Green Tringe S, Barry K, Malfatti S, Goessmann A, Robinson-Rechavi M, van der Meer JR. 2013. Community-wide plasmid gene mobilization and selection. *ISME J* **7**(6): 1173–1186. doi: 10.1038/ismej.2013.13

Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, Yousef M, Watson AR, Lolans K, Esen ÖC, et al. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol* **21**(1): 292. doi: 10.1186/s13059-020-02195-w

Shannon KE, Lee D-Y, Trevors JT, Beaudette LA. 2007. Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Science of The Total Environment* **382**(1): 121–129. doi: https://doi.org/10.1016/j.scitotenv.2007.02.039

Sheik CS, Beasley WH, Elshahed MS, Zhou X, Luo Y, Krumholz LR. 2011. Effect of warming and drought on grassland microbial communities. *ISME J* **5**(10): 1692–1700. doi: 10.1038/ismej.2011.32

Sheppard RJ, Beddis AE, Barraclough TG. 2020. The role of hosts, plasmids and environment in determining plasmid transfer rates: A meta-analysis. *Plasmid* **108**: 102489. doi: https://doi.org/10.1016/j.plasmid.2020.102489

Shinoda M, Nachinshonhor GU, Nemoto M. 2010. Impact of drought on vegetation dynamics of the Mongolian steppe: A field experiment. *J Arid Environ*, in press. doi: 10.1016/j.jaridenv.2009.07.004

Shintani M, Sanchez ZK, Kimbara K. 2015. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* **6**: 242. Switzerland. doi: 10.3389/fmicb.2015.00242

Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, Fernandez-Guerra A, Liebeke M, Schweder T, Polz MF, et al. 2020. Verrucomicrobia use hundreds of enzymes

to digest the algal polysaccharide fucoidan. *Nat Microbiol* **5**(8): 1026–1039. doi: 10.1038/s41564-020-0720-2
Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**(5): 865–891. Blackwell Publishing Ltd Oxford, UK.
Smalla K, Jechalke S, Top EM. 2015. Plasmid Detection, Characterization, and Ecology. *Microbiol Spectr* **3**(1): PLAS-0038-2014. doi: 10.1128/microbiolspec.PLAS-0038-2014
Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. 2010. Mobility of plasmids. *Microbiol Mol Biol Rev* **74**(3): 434–452. United States. doi: 10.1128/MMBR.00020-10
Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**(7376): 241–244. doi: 10.1038/nature10571
Stalder T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. *ISME J* **13**(10): 2437–2446. doi: 10.1038/s41396-019-0446-4
Stasiak G, Mazur A, Wielbo J, Marczak M, Żebracki K, Koper P, Skorupska A. 2014. Functional relationships between plasmids and their significance for metabolism and symbiotic performance of Rhizobium leguminosarum bv. trifolii. *J Appl Genet* **55**(4): 515–527. doi: 10.1007/s13353-014-0220-2
Stecher B, Denzler R, Maier L, Bernet F, Sanders MJ, Pickard DJ, Barthel M, Westendorf AM, Krogfelt KA, Walker AW, et al. 2012. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences* **109**(4): 1269 LP – 1274. doi: 10.1073/pnas.1113246109
Strickland MS, Lauber C, Fierer N, Bradford MA. 2009. Testing the functional significance of microbial community composition. *Ecology*, in press. doi: 10.1890/08-0296.1
Stuart YE, Campbell TS, Hohenlohe PA, Reynolds RG, Revell LJ, Losos JB. 2014. Rapid evolution of a native species following invasion by a congener. *Science (1979)* **346**(6208): 463–466. American Association for the Advancement of Science.
Stylinski CD, Allen EB. 1999. Lack of Native Species Recovery Following Severe Exotic Disturbance in Southern Californian Shrublands. *Journal of Applied Ecology* **36**(4): 544–554. [British Ecological Society, Wiley].
Svara F, Rankin DJ. 2011. The evolution of plasmid-carried antibiotic resistance. *BMC Evol Biol* **11**(1): 130. doi: 10.1186/1471-2148-11-130
Swift MJ, Heal OW, Anderson JM. 1979. Decomposition in terrestrial ecosystems. *Studies in Ecology*, in press.
Thomas CM, Nielsen KM. 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Microbiol* **3**(9): 711–721. doi: 10.1038/nrmicro1234
Tintjer T, Rudgers JA. 2006. Grass-herbivore interactions altered by strains of a native endophyte. *New Phytologist*, in press. doi: 10.1111/j.1469-8137.2006.01720.x
Touchon M, Perrin A, de Sousa JAM, Vangchhia B, Burn S, O'Brien CL, Denamur E, Gordon D, Rocha EPC. 2020. Phylogenetic background and habitat drive the genetic diversification of Escherichia coli. *PLoS Genet* **16**(6): e1008866. Public Library of Science.
Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol* **6**: 771. Switzerland. doi: 10.3389/fmicb.2015.00771

Tu Q, Lin L, Cheng L, Deng Y, He Z. 2019. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics* **35**(6): 1040–1048. doi: 10.1093/bioinformatics/bty741

Ulrich DEM, Sevanto S, Ryan M, Albright MBN, Johansen RB, Dunbar JM. 2019. Plant-microbe interactions before drought influence plant physiological responses to subsequent severe drought. *Sci Rep*, in press. doi: 10.1038/s41598-018-36971-3

UNITE Community. 2019. UNITE QIIME release for Fungi. *UNITE Community*, in press. doi: 10.15156/BIO/786334

Vilà M, Burriel JA, Pino J, Chamizo J, Llach E, Porterias M, Vives M. 2003. Association between Opuntia species invasion and changes in land-cover in the Mediterranean region. *Glob Chang Biol* **9**(8): 1234–1239. John Wiley & Sons, Ltd (10.1111). doi: 10.1046/j.1365-2486.2003.00652.x

Wardle DA, Yeates GW, Barker GM, Bonner KI. 2006. The influence of plant litter diversity on decomposer abundance and diversity. *Soil Biol Biochem* **38**(5): 1052–1062. doi: https://doi.org/10.1016/j.soilbio.2005.09.003

Warton DI, Wright ST, Wang Y. 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol Evol*, in press. doi: 10.1111/j.2041-210X.2011.00127.x

Whelan S, Goldman N. 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol* **18**(5): 691–699. doi: 10.1093/oxfordjournals.molbev.a003851

Wick Ryan R, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**(6): e1005595. Public Library of Science.

Wick Ryan R., Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* **3**(10): e000132. Microbiology Society. doi: 10.1099/mgen.0.000132

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**(20): 3350–3352. Oxford University Press. doi: 10.1093/bioinformatics/btv383

Wickham H. 2009. *Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.* Media. doi: 10.1007/978-0-387-98141-3

Wollman EL, Jacob F. 1956. Sexuality in bacteria. *Sci Am* **195**(1): 109–119. Scientific American, a division of Nature America, Inc.

Wood JL, Tang C, Franks AE. 2018. Competitive traits are more important than stress-tolerance traits in a cadmium-contaminated rhizosphere: A role for trait theory in microbial ecology. *Front Microbiol*, in press. doi: 10.3389/fmicb.2018.00121

Xu Y, Lewandowski K, Lumley S, Pullan S, Vipond R, Carroll M, Foster D, Matthews PC, Peto T, Crook D. 2018. Detection of Viral Pathogens With Multiplex Nanopore MinION Sequencing: Be Careful With Cross-Talk . Frontiers in Microbiology .

Xue H, Cordero OX, Camas FM, Trimble W, Meyer F, Guglielmini J, Rocha EPC, Polz MF. 2015. Eco-Evolutionary Dynamics of Episomes among Ecologically Cohesive Bacterial Populations. *mBio* **6**(3): e00552-15. doi: 10.1128/mBio.00552-15

Yang Y, Li B, Ju F, Zhang T. 2013. Exploring Variation of Antibiotic Resistance Genes in Activated Sludge over a Four-Year Period through a Metagenomic Approach. *Environ Sci Technol* **47**(18): 10197–10205. American Chemical Society. doi: 10.1021/es4017365

Yoon J-H, Wang S-YS, Gillies RR, Kravitz B, Hipps L, Rasch PJ. 2015. Increasing water cycle extremes in California and in relation to ENSO cycle under global warming. *Nat Commun* **6**(1): 8657. doi: 10.1038/ncomms9657

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**(W1): W95–W101. England. doi: 10.1093/nar/gky418

Zhao Y, Yang B, Li M, Xiao R, Rao K, Wang J, Zhang T, Guo J. 2019. Community composition, structure and productivity in response to nitrogen and phosphorus additions in a temperate meadow. *Science of the Total Environment*, in press. doi: 10.1016/j.scitotenv.2018.11.155